Implementation of new assistive technologies for people affected by Autistic Spectrum Disorders (ASDs)

PhD Thesis

PhD course on Automatic, Robotics and Bioengineering

XXIII (2008-2010)

SSD: ING-INF/06

Student: Antonino Armato

Supervisors: Prof. Danilo De Rossi, Prof. Luca Fanucci

UNIVERSITÀ DI PISA

University of Pisa

Interdepartmental Research Center "E.Piaggio"

Pisa March 20, 2011

I would like to dedicate this work to Stefania
because I can do nothing without her love.
She had my love right from the start and
I know she always will.

Acknowledgements

I am grateful to my supervisors who have provided great academic suggestions for my research.
I would like to thank my colleagues and friends of the BIOLAB: Enzo Pasquale Scilingo, Gaetano Valenza for their help in setting up experiments, looking at data, and providing fruitful advices. Special thanks to Antonio Lanata', with him I shared ideas, difficulties. His contribution in my research has been fundamental thanks his great technical skills which in many ways inspired me.
I am grateful to my girlfriend and my family who supported and encourage me during the completion of my project.

Abstract

Individuals with Autistic Spectrum Disorders (ASDs) have impairments in the processing of social and emotional information. The number of children known to have autism has increased dramatically since the 1980s. This has sensitized the scientific community to the design and development of technologies suitable for treating an autistic patient in order to broaden the emotive responsiveness, such as the employment of robotic systems to engage proactive interactive responses in children with ASDs.

My PhD work focuses on the design and develop of new technologies for therapy with individual affect by ASD. The main challenge of my work has been to design and develop a novel control architecture able to reproduce the brain characteristics in terms of high concurrency processing, flexibility and the ability to learn new behavior. The main difficulties in implementing Artificial Neural Networks (ANNs) in hardware in terms of accuracy, gate complexity and speed performance are discussed.

A new wearable eye tracking system able to investigate attention disorders early in infancy is proposed. Technological choices are emphasized with respect to unobtrusive and ecological to adapt the device for infants. New algorithms to increase the system robustness under illumination change and during calibration process have been developed and herein presented. Experimental test prove the effectiveness of the solutions.

Considerations on the future research directions are addressed, stressing the multiple application fields of the designed device.

Contents

1 Introduction **1**

2 Background and Motivations **5**
2.1 Historical development of autism 5
2.2 Diagnosis and intervention 15

3 Assistive Technology **23**
3.1 Introduction 23
3.2 Educational Software 25
3.2.1 Fun with Feelings 25
3.2.2 Bubble Dialogue 26
3.2.3 Mental state simulator 27
3.2.4 Mind-reading 28
3.3 Virtual Environments 30
3.4 Robotics 32
3.4.1 Robots in Aurora Project 33
3.4.2 Storytelling robots 38
3.4.3 Infanoid and Keepon 38
3.4.4 Paro 44

3.5 Wearable Devices 45

4 FACE, an embodied interactive social interface 49
4.1 Face Robot Hardware 53
4.2 The cognitive architecture 54
4.3 Neural Networks in Hardware 56
4.3.1 State of the Arts 57
4.3.2 A new approximation method 60
4.3.3 Sigmoid approximation 65
4.3.4 Hyperbolic tangent approximation 72
4.3.5 Hyperbolic tangent by the sigmoid 74
4.3.6 VLSI Circuit Design 79
4.3.7 Comparative evaluation 87
4.3.8 Topology of the ANNs in Hardware 91
4.4 Results and Discussions 94

5 Eye Tracking System 97
5.1 State of the arts 99
5.2 HATCAM 107
5.3 Eye tracking Method 109
5.3.1 Eye Extraction 112
5.4 Photometric normalization techniques 113
5.4.1 Pupil tracking 119
5.4.2 Ellipse fitting 121
5.4.3 Mapping of the eye positions in the scene . . . 124
5.4.4 Correction process 125
5.4.5 Vestibulo-ocular calibration curve 132
5.5 Determination of VOR curve 132

5.6 Experimental study of photonormalization technique 135
5.7 Experimental validation of correction process 141
5.8 Eye tracking as response to affective stimuli 153
5.8.1 Post-processing and feature extraction 155
5.8.2 Classification 158
5.8.3 Classification Results 160
5.9 Results and Discussions 163

6 Conclusions and Future Works **167**

Bibliography **176**

List of Figures

3.1 Fun with feelings. 26
3.2 Screenshots from Mind Reading 29
3.3 Therapist User Interface 31
3.4 KASPAR robot 34
3.5 A child engages in a tactile exploration of KASPAR . 35
3.6 Robota robot 37
3.7 Storellying robot 39
3.8 Infanoid robot 41
3.9 Keepon robot 43
3.10 Paro robot 44
3.11 Wearcam 48

4.1 FACE Carrara 50
4.2 FACET platform 52
4.3 FACE Robot 53
4.4 Control system 55
4.5 Approximation with area method 63
4.6 Maximum of the relative error in function of c 67
4.7 Approximated sigmoid function 68
4.8 Relative error of the sigmoid function 69

4.9 Approximated derivative sigmoid function 71
4.10 Approximated hyperbolic tangent with areas method 75
4.11 Absolute error of hyperbolic tangent with areas method 76
4.12 Hyperbolic tangent function with areas method . . . 78
4.13 Approximated hyperbolic tangent 80
4.14 Absolute error of hyperbolic tangent 81
4.15 Approximated derivative hyperbolic tangent 82
4.16 Average convergence of network using FXP formats . 84
4.17 Activation Function-Scheme 1. 86
4.18 Activation Function-Scheme 2. 87
4.19 Generator of Hyperbolic Tangent-Scheme 3 88
4.20 Comparison frequency vs error 89
4.21 Comparison logic element vs error 90
4.22 The Virtual Reconfigurable Network (VRN) 92
4.23 Convergence of different species 93
4.24 Number of individuals for different species 93

5.1 Contact lens system 100
5.2 Pelz system 101
5.3 Pelz system 102
5.4 IR illumination 103
5.5 Infant wearing the eye-tracker 106
5.6 Version I: "Baseball-like" hat configuration. 107
5.7 Version II: "Head band" configuration. 108
5.8 A control subject wear HATCAM. 108
5.9 Representation of the axes 110
5.10 Block diagram of the EGT processing 111
5.11 Example of a single frame captured by the camera . . 112

LIST OF FIGURES

5.12 Red component of the ROI. 113
5.13 Block diagram of the pupil tracking algorithm 121
5.14 Example of the histogram of the eye 122
5.15 Example the eye image after the binarization process. 122
5.16 Application of the pupil tracking algorithm 123
5.17 Block Diagram of the mapping function 125
5.18 Representation of the axes 127
5.19 Calibration plane. 128
5.20 Initial condition of the calibration. 128
5.21 Representation of the positive θ_h rotation. 129
5.22 Final condition after the positive θ_h rotation 129
5.23 Experimental VOR calibration (θ) 133
5.24 Experimental VOR calibration (ψ) 133
5.25 Average gain around vertical axis 134
5.26 Average gain around horizontal axis 134
5.27 Example of the points showed on the screen 136
5.28 Example of all photonormalization techniques 137
5.29 Example of the distributions of errors 138
5.30 Subject emplacement 145
5.31 Statistical comparison along x with VOR 152
5.32 Statistical comparison along y with VOR 152
5.33 Gaze points during neutral elicitation 160
5.34 Gaze points during arousal elicitation 161

6.1 FACET platform 168
6.2 Version I: "Baseball-like" hat configuration. 169
6.3 Version II: "Head band" configuration. 169
6.4 A control subject wear HATCAM. 172

List of Tables

4.1 Polynomial approximation of the sigmoid function . . 66
4.2 Summary of absolute and relative errors 70
4.3 Polynomial approximation of the s. derivative 72
4.4 Summary of absolute errors in the published literature 73
4.5 Polynomial approximation of the hyperbolic tangent . 74
4.6 Polynomial approximation of the h.t.derivative 77
4.7 Error in function of the precision bits 85

5.1 Average Gain VOR 135
5.2 Median and dispersion of eye tracking errors in pixels 142
5.3 Median and dispersion of pupillometry errors in $pixel^2$ 142
5.4 Median and dispersion of the execution time [s] . . . 142
5.5 Eye tracking error classes 143
5.6 Pupillometry error classes 144
5.7 X accuracy with head in stationary mode 147
5.8 Y accuracy with head in stationary mode 148
5.9 X accuracy with the head free to move 149
5.10 Y accuracy with the head free to move 150
5.11 Kruskal-Wallis test:stationary vs movement 151
5.12 Execution time 151

5.13 Features extracted from RQA 162
5.14 Confusion Matrix of K-NN classifier 162

Chapter 1

Introduction

Individuals with Autistic Spectrum Disorders (ASDs) have impairments in the processing of social and emotional information. The number of children known to have autism has increased dramatically since the 1980s. The most recent reviews of epidemiology estimate a prevalence of about to 6 per 1000 for ASD. This has sensitized the scientific community to the design and development of technologies suitable for treating an autistic patient in order to broaden the emotive responsiveness, such as the employment of robotic systems to engage proactive interactive responses in children with ASDs. Understanding and teaching the processing of socio-emotional abilities is the inspiring principle of this novel approach and is believed to be of tremendous clinical significance. Encouraging studies with robotic dolls, mobile robots and humanoids acting as social mediators have provided important insights and demonstrate the necessity of long term studies.

An android-based treatment focused on the embodiment of emo-

tions, empathy and interactive imitation is herein proposed. This treatment was carried out at the Interdepartimental Research Center "E. Piaggio." The hypothesis is that adaptive therapy using a facial robot named FACE (Facial Automaton for Conveying Emotions) endowed with the ability to sense, adapt and respond to a patient's postulated emotional and mental states will enable autistic subjects to learn empathy and gradually enhance their social competence. The integrated sensing, monitoring, processing and emotionally responsive android-based therapeutic platform is termed FACET (FACE Therapy). Physiological signals (HR, HRV, RR, EMG and EDA) and eye gaze are acquired using a comfortable, unobtrusive sensorized shirt and a cap containing integrated cameras and mini gyroscopes (eye tracking system). The complexity of the input signal management in terms of patient emotion classification, synchronization signals and motor control requires a complex hardware system able to learn and process all of the inputs in real-time. My PhD work focuses on the design and develop of new technologies for therapy with individual affect by ASD. The main challenge of my work has been to design and develop a novel control architecture able to reproduce the brain characteristics in terms of high concurrency processing, flexibility and the ability to learn new behavior. The main difficulties in implementing Artificial Neural Networks (ANNs) in hardware in terms of accuracy, gate complexity and speed performance will be discussed. Another challenge of the present work is to provide a wearable eye tracking system able to investigate attention disorders early in infancy. The hardware solution suited for infants in terms of unobtrusiveness and aesthetics will also be discussed. This thesis dissertation is structured as follows:

- In Chapter 2, an overview of autism is presented.
- In Chapter 3, a description of the assistive technology in autism is presented, and motivations will be explained as to why people with autism generally feel comfortable in predictable environments, and especially enjoy interacting with computers and robots.
- In Chapter 4 the FACE-T platform is described. A digital implementation of Artificial Neural Networks (ANNS) will be described in detail. A new method of piecewise line approximation will be introduced from a mathematical point of the view. Moreover, a complete digital implementation study will be discussed according to accuracy, gate complexity and speed performance. A convergence study will be taken into account. A flexible topology of the neural networks is proposed.
- In Chapter 5 a deep description of a new eye tracking system is presented. The software processing is described along with the hardware solution. Photometric techniques will be proposed to increase the robustness of the device under illumination change. A new algorithm to overcome the main drawback of the calibration process is realized. Experimental validation sections will be showed.
- In Chapter 6 the main outcomes of this work will be summarized and future research directions will be addressed, stressing the multiple application fields of the designed device.

Chapter 2

Background and Motivations

2.1 Historical development of autism

The term autism (from the greek *auto's*:"himself") was coined by Swiss doctor E. Bleuler (1911) [1] to describe a symptom of "dementia praecox", referring to the behavior of some individuals during childhood. This symptom is characterized by a "narrowing of relationships with persons and with the outside world, a narrowing so extreme as to exclude anything but your own self." Cases of precocious dementia "had been previously described by De Sanctis (1905) and Heller (1909). The latter makes reference to dementia during childhood, "Heller's dementia, which occurs between 3-4 years of age and is slow in running its course, resulting in the loss of mental functions and speech. Often intelligent facial expressions and psychomotor excitement are observed. A few years later, De Sanctis describes

another syndrome that is similar to "Heller dementia" named "late aparetico-aphasic." Early and rapid onset are the forms described by Cerberus (1926), calling this "regression mentis-infanto juvenilis". In this case, a histological study with brain biopsy resulted in the discovery of lipid degeneration. Initially, autism was classified as a disorder of schizophrenia, a Bleuleran term used in 1916, that replaced "dementia praecox". This initial reference to schizophrenia proved to be misleading. In fact, it gave little consideration to the fact that the symptom of "autism" in schizophrenic patients occurs after a short or long period of prosperity even extending into adulthood, and not as a modification of previous normal relations with the world . In children, however, autistic symptoms are not only very early, but it is often impossible to identify their previous state of wellness.

The merit of defining autism as a unitary and separate pathological entity belongs to the American psychiatrist Kanner L. In an article published in 1943 in the journal "Nervous Child" [2], he reported an analysis conducted for 5 years on 11 children who reveal, from early childhood, an inability to develop normal affective contacts. He writes, "we have to assume that these children have come into the world with an innate inability to give rise to normal affective contact, biologically provided to persons, as well as other children have come into the world with innate physical or intellectual handicap" Kanner is sure to be confronted with "pure cases of autistic congenital troubles of affective contact". For the first time, autism is defined as a separate "syndrome" (early infantile autism), having evolutionary and congenital peculiarities with nine key features:

2.1 Historical development of autism

1. peculiarities in social relations;
2. speech disorder;
3. good memory and learning;
4. eating disorders;
5. excessive emotional reactions;
6. adherence to routines;
7. good relationships with physical objects;
8. motor awkwardness;
9. born of intellectual parents.

Almost in the same period in 1944, the Austrian doctor H. Asperger published a study of a group of persons with a disorder he called "autistic psychopathy" [3], characterized by social isolation, stereotypes, and resistance to changes in routine. The symptoms are similar to those of the 11 patients of Kanner who, however, presented either with total absence of word use (3 cases) or severely delayed language development. Asperger's syndrome differs from autistic disorder precisely due to the lack of specific language disorders. The two psychiatrists, independently, use the term autism, perhaps borrowing from Bleuler, but in science it was discussed whether the two disorders were distinct or overlapping. The most accepted thesis is that Kanner and Asperger's definition of autism makes up part of the same spectrum, with the first part at the lower end (low-functioning subjects) and the second part at the upper end (high-functioning

individuals). Kanner, following his observations of the parents of his 11 patients, who were too busy intellectually, perfectionists, with little sense of humor, was convinced that the cause of autism could be found in affective deprivation, giving strength to the psychoanalytic theories. In an article published in Time Magazine in 1960, Kanner said, "In the whole group there were very few loving fathers and mothers. Children with early infantile autism were born to parents particularly efficient, professional, cool and rational, who "thawed" only for the time to produce a child. B. Bettelheim embraced the idea of Kanner, setting his work on psychogenic causes of autism, focusing on the origin of the disorder in early mother-child relations. A distorted relation is the owing of autism. In 1967, Bettelheim published "empty fortress" [4], the synthesis of his work at the University of Chicago, which coined the term "refrigerator mother" to describe the coldness and detachment with which the mothers of autistic children deal with children. The quote, below, shows clearly his model: "In dealing with the origin of the extreme situations in early childhood, we can now say that the pathology of the mother is often very severe and in many cases her behavior toward the child is a particularly significant example of an abnormal relationship . In part, the mother, because of frustration in maternal feelings or because of the child's anxiety, can act, instead of with sweet insistence, with anger or indifference. This, in turn, lends itself to creating new anxiety in the child. Any refusal tends to weaken the impulse of the newborn to observe the surrounding environment and act upon it, and without such impetus, personality can not develop." The spread of this devastating ideology created misunderstandings between parents and professionals and even to-

day, theories that consider autism as a result of maternal-affective deficiencies are widespread among the public and professionals and continue to produce needless suffering in children and parents that could otherwise easily be avoided. Indeed, Bettelheim's therapeutic approach consists of removing the children from the mother and administering psychotherapeutic treatment to the mothers or parents, as if the healing of the child could indirectly depend on these treatments. Their placement with other families, thought to be safer on the affective and relational levels, certainly did not eliminate the autism. Over the years, many elements did not confirm the origin of psychogenetic autism. Instead, it was recognized that the early onset of autistic symptoms derived from a biological neurocerebral dysfunction. Meanwhile, in the sixties, America and Europe created the first Centers and the first Society dedicated to a systematic study of autism. These include, the National Society for Autistic Children and the Autism Society of America. The need to specify diagnostic criteria marked the birth of the Diagnostic and Statistical Manual of Mental Disorders (DSM), which, in its second edition [5], classifies autism as a childhood schizophrenia, using synonyms for "childhood psychosis" and "symbiotic psychosis." In 1969, at the first meeting of the National Society for Autistic Children, Kanner, returning to his first case of autism as a disorder of natural development, said, "I still know little about autism. I know little just because there is still much to discover. We must be prudent. we seek to acquire new knowledge with a weighted curiosity. We have to follow different tracks to improve our knowledge. We need to test various theories about possible causes, and exonerate you, the parents. " In 1970, the international magazine on autism was founded,

The Journal of Autism and Childhood Schizophrenia, now called the Journal of Autism and Developmental Disorders. In France, the scientific bulletin ARAP (Association pour la Recherche sur l'Autisme et la Psychose Infantile) changed its name to the Association pour la Recherche sur l'Autisme et la Prevention des Inadaptations. However, in 1972, Niko Tinbergen and Elisabeth [6] rejuvenated the older theories. Autism main origins is in the deficiency of the normal development in the link between mother and child. The scientific evidence contrary to this thesis did not discourage Tinbergen, whose thesis provides the basis for their care practice (holding) of autism. The consequences of holding would prove to be completely negative and severely criticized by Frith (1984) [7] [8] and L. Wing (1986) [9] and was quickly abandoned. More specifically, Wing defines autism as a developmental disorder in which the main characteristics are related to socialization and language problems combined with the lack of interest and imagination. In the meantime, in the United States of America, E. Schopler founded the TEACCH program (Treatment and Education of The Autistic and Related Communication Handicapped Children), which consists of a set of services that deal with autistic patients and their families from diagnosis to adulthood. The program is based on the collaboration between an educational team and family along with the training of operators. In 1980, in the third edition of the DSM manual (DSM III) [10], autism was no longer considered a psychosis, but was included among the Pervasive Developmental Disorders.

V.D. Sanua, in an article published in 1987, using correct statistical methods and objective information on the families of autistic children, definitively rejected the idea of the "refrigerator-mother."

In the same year, the tenth edition of the Diagnostic Manual of Mental Disorders (ICD-10)[11] was published by the World Health Organization that, simular to the DSM III, classifies autism in the Pervasive Developmental Disorders. Meanwhile, the development in neuroscience with the possibility to investigate the influence of genetics and immunology, extends the knowledge of neurophysiological processes and mental disabilities, and, in particular, autism. It would be simplistic to seek solutions to autism as conceiving man only in his psychic, emotional and affective fields, while ignoring his biological dimension. In this context of neuro-psycho-organic unity, autistic pathology is placed.

Meanwhile, the fourth revised edition of DSM (1994) [12], maintains the diagnostic criteria for autism under the heading of Pervasive Developmental Disorders. In the early nineties, G. Rizzolati [13] and colleagues at the University of Parma, discovered a class of neurons located in the premotor cortex of monkeys that became activated when the animal performs an action or when he sees her make by the researcher.

Subsequent studies by L. Fadiga et al. (1995) [14], J. Decety et al. (1997) [15], and M. Iacoboni et al. (1999) [16] identified a similar system of mirror neurons in humans. Using laboratory techniques, they understood that mirror neurons have the capacity to process the action of others, "as if" the viewer was directly the one who is performing. This ability has a significant role in imitation, language, empathy and behavior, and in autism, the deficits coincide precisely with the supposed function of mirror neurons.

Since 2000, many laboratories, inspired by this premise, have explored the possibility of the dysfunction of mirror neurons in autis-

tic individuals. The first study centered on the μ electroencephalographic wave, which is inactive in normal individuals who do manual activity or who only observe It is rather sensitive to the execution of actions but not sensitive to their observation in individuals with autism. Following studies using the magnetoencephalography technique (MEG), Magnetic Resonance Imaging (MRI) and functional magnetic resonance imaging (fRMI) individuals with autism showed a dysfunctional mirror neuron system. In any case, it must be considered that a brain does not work in isolation independently of psychological processes. Most likely, the mirror neuron system responds to all human actions in neurotypical individuals because these subjects are able to relate to everything, while the autistic person responds to a narrow range of stimuli. A recent experiment conducted on individuals with both autism and neurotypical by fRMI, based on the observation of simple hand movements, has reported that the two groups showed a similar activation of mirror neuron areas. The theory seemed to falter, but Iacoboni immediately pointed out that in the more complex tasks such as those related to these experiments, the results would certainly be different for subjects with autism because the deficit of mirror neurons could cause a lack of development or a partial development of socio-behavioral and language skills.
In summary, if the behavioral symptoms of autism are related to a dysfunction of the mirror neuron system, it is possible to propose a "paradigm training through neurofeedback" in order to improve its functionality and the behavioral abnormalities caused by this dysfunction. Finally, it is worth noting the studies of J. Pineda [17], which aimed to demonstrate that the dysfunction of mirror neurons are the result of defects in neuronal connectivity. Oberman et al.

[18] have carried out research based on Transcranial Magnetic Stimulation (TMS) to evaluate the possibility of potentially changing the patterns of neural connection.
Autism has no geographical and/or ethnic prevalence because it has been described in all populations in the world, in every race and social background. Based on currently available data, the prevalence of the disease is passed from 2-5:10.000 born before 1985 in 10-20:10.000 since 2000. Several recent studies in North America and Europe report an incidence of the disease approximately 1:150 births. The syndrome appears to affect males in a measure 3-4 times higher than females. According to most researchers, autism is much more frequent than 30 years ago. This discrepancy in the estimates of prevalence of the syndrome over time could be attributed to a number of factors including a new definition of the diagnostic criteria rather than a real increase in cases, and with the inclusion of milder forms as well as the diffusion of standardized diagnostic procedures.
Regarding the link between autism and genetic factors, many family studies confirm the importance of heredity. For example, the probability that two homozygous twins are both autistic is much higher than for two heterozygote twins. Also, a couple with an autistic child, in 2-8 % of cases will have a second child with autism. Frequently, some family members of an autistic person present characteristics similar to autism but in a milder form. Between autism and mental retardation there is a very narrow relationship (from 60% to 90%, mostly of medium or severe grade) as well as between autism and epilepsy (50% of cases). The epidemiological findings mentioned above have prompted research groups to identify the genes involved

in autism. The strongest evidence that has emerged from this research is that there is no one "gene" of autism, but rather a number of genes that contribute to confer a vulnerability to the onset of the disorder (Bailey et al. 1996; Szatmri et al. 1998; Folstein et al., 2001). The gene loci of interest were identified on chromosome 7 (IMGSAC, 1998, CLSA, 1999; IMGSAC, 2001a), on chromosome 2, 16 and on chromosome 17 (IMGSAC, 2001b). On this direction, the genes involved can be many and diverse in nature. Therefore, the rigid paradigm of "one-gene-one-disorder" needs to be revised. In fact, in the complicated development process of the brain, a series of genes with diversified functions are coordinated (activation, modulation, inhibition). From these interactions, the morpho-functional traits necessary are realized for their use the experiential data and their organization in the knowledge system and relationship systems. Recently, in the journal "Nature Genetics" in 2007, the Consortium Autism Genome Project, which consists of more than 50 research centers in the U.S. and Europe, published a study reporting the results of a survey of about 1200 families with at least two cases of autism in order to assess genetic variants linked to the syndrome. With the technique of DNA array possible areas involved in autism have been identified, particularly in the short beam of chromosome 11. In one case (1 / 1200) a gene defect of neurexina 1 was identified. The data that appears is scarcely significant from a statistical point of view. This is very interesting because this protein regulates the contacts and communications between neurons. Clearly, the results may not be used for diagnostic tests. However, the Consortium hopes to identify, with subsequent studies and detailed analysis of DNA using higher resolution technique, other genes related to the

disease, correlating gene defects with patient symptoms. The multifactorial nature of autism could involve several genetic factors in different ways, whose identification will be long and laborious [19]. Although there is an increased interest in the relationship between autism and autoimmune diseases, actually there is no evidence that immunological mechanisms may cause or contribute to the emergence of organic abnormalities found in autism. In 1998, Wakefield and others put forward the hypothesis of a temporal relationship between vaccinations and the occurrence of certain autistic behaviors, but at present there is no data indicating that a vaccine increases the risk of developing autism or any other behavioral disorder (Parker et al., 2004).

2.2 Diagnosis and intervention

The historical development of autism clearly highlights the multifactorial nature of the syndrome. The wide variability of situations in which the autistic disorder may occur requires an equally important heterogeneity of the severity of symptoms and mental limitations. For this reason, the term "autistic spectrum" is frequently used. The diagnosis depends on the recognition of complex behavioral patterns that have been observed over time and in the current clinical description. Therefore, to avoid heterogeneity and variability, which could generate confusion and the dispersion of ideas, it is essential for the operators to have fixed international references. Currently, the international classification ICD 10 [10] and DSM IV [11] are widely used, which place autism among the "Pervasive Devel-

opmental Disorders." The following, definition of Autism in the tenth edition of ICD is as follows: "A pervasive developmental disorder defined by the presence of abnormal and/or impaired development that is manifest before the age of 3 years, and by the characteristic type of abnormal functioning in all three areas of social interaction, communication, and restricted, repetitive behaviour. The disorder occurs in boys three to four times more often than in girls". The diagnostic criteria of DSM IV is reported:
" **(I)** A total of six (or more) items from (A), (B), and (C), with at least two from (A), and one each from (B) and (C).
(A) qualitative impairment in social interaction, as manifested by at least two of the following:

1. marked impairments in the use of multiple nonverbal behaviors such as eye-to-eye gaze, facial expression, body posture, and gestures to regulate social interaction;

2. failure to develop peer relationships appropriate to developmental level;

3. a lack of spontaneous seeking to share enjoyment, interests, or achievements with other people, (e.g., by a lack of showing, bringing, or pointing out objects of interest to other people);

4. lack of social or emotional reciprocity (note: in the description, it gives the following as examples: not actively participating in simple social play or games, preferring solitary activities, or involving others in activities only as tools or "mechanical" aids).

(B) qualitative impairments in communication as manifested by at least one of the following:

1. delay in, or total lack of, the development of spoken language (not accompanied by an attempt to compensate through alternative modes of communication such as gesture or mime);

2. in individuals with adequate speech, marked impairment in the ability to initiate or sustain a conversation with others;

3. stereotyped and repetitive use of language or idiosyncratic language;

4. lack of varied, spontaneous make-believe play or social imitative play appropriate to developmental level.

(C) restricted repetitive and stereotyped patterns of behavior, interests and activities, as manifested by at least two of the following:

1. encompassing preoccupation with one or more stereotyped and restricted patterns of interest that is abnormal either in intensity or focus;

2. apparently inflexible adherence to specific, nonfunctional routines or rituals;

3. stereotyped and repetitive motor mannerisms (e.g hand or finger flapping or twisting, or complex whole-body movements);

4. persistent preoccupation with parts of objects.

(II) Delays or abnormal functioning in at least one of the following areas, with onset prior to age 3 years: (A) social interaction.
(B) language as used in social communication.
(C) symbolic or imaginative play.
(III) The disturbance is not better accounted for by Rett's Disorder or Childhood Disintegrative Disorder".

These diagnostic criteria indicate the complexity and pervasiveness of disorders that early and significantly undermine the normal development of the child, with its appearance before 3 years of life and having the following functional impairment:

- social interaction;
- communication;
- interests and activities.

Therapeutic treatment includes the following actions:

- improve social interaction;
- enhance communication;
- promote widening interest and greater flexibility in patterns of action.

This clarification, which may seem obvious, intends to underline the need for consistency; if a clinical-descriptive approach is taken into account, the planning of interventions must be based on an equally clear definition of the objectives. The strategies commonly

recommended and adopted, although varies in relation to a number of factors such as age or degree of functional impairment, may be classified into two broad categories:

- behavioral approaches;
- evolutionary approaches.

In this perspective, the "behavioral analyst" uses the data obtained to formulate theories related to determinate behavior that occurs in a particular context (Behavior Analysis). Consequently, the analyst implements a series of actions designed to change behavior and/or context. The information obtained will be used proactively and systematically to change the behavior. The Applied Behavior Analysis takes into account the following four elements:

- the background (everything that anticipates the behavior in question);
- the conduct taken into account (which must be observable and measurable);
- the consequences (which is derived from the conduct in question);
- the context (defined in terms of location, people, materials, activities or time of day) in which the behavior occurs.

The intervention program is carried out on data emerging from the analysis, using the usual techniques of behavior therapy:

- prompting;

- fading;
- modeling;
- shaping and reinforcement.

Concerning the evolutionary (or interactive) approaches, for the purpose of this work, it is more interesting to study the emotional and relational dimension where the child is acting. Normally, the different areas of emotions, cognitive function and communication skills that evolve and influence each other define a dynamic system that cannot be considered the sum of its components involved in the entire realization.
Practically, we have an open dynamic system that, in relation to experiential contributions, is evolving to more and more advanced functional levels without the identification of the changes of individual components, which is crucial. In this perspective the action is characterized as a "child-centered" action to encourage free expression, initiative, and participation. The environment is not only conceived as a physical space in which intervention programs are implemented but assumes a "therapeutic" value because of the environment of interaction.
In the following section, we will show how biomimetic robots and androids play an important role in evolutionary approaches (interactive) underling as in autistic individual, it is possible simplify the non verbal communication levels using facial expressions, vocalizations, prosody, mimic, and posture. Recent research has shown that certain subjects with cognitive deficits perceive and treat robots not as machines, but as their artificial partners [20] [21] and that chil-

dren with autism are interested in playing with mechanical toys or computers. Social and emotional reactions may be transmitted to individuals with ASD via three-dimensional displays based on social or cybot robot [22] [23], thereby stimulating the curiosity of the child, who understands that he is before a machine, but who considers the machine as a human. The goal is to establish a clinical innovative protocol in the assessment and treatment of autism going about the interactive approach based on imitation and a pragmatic use of social and emotional reciprocity. In conclusion, we can say that, at present, there is no cure to "healing" the autistic individual, but it is possible to activate a series of interventions that can lead to substantial improvements. The placement of the specificity of problems presented and a particular methodological attention in the standard procedures tend to bring the patient to obtain maximum independence and an increase in social integration. In activities and research, the concept of "quality of life" must also be taken into consideration, both of the children and their families, as well as "family weight" which is meant as an emotional expenditure of energy, as well as relational and financial expenditures that are necessary to address in taking care of the little patient. Lastly, the various services related to autism must provide support to address the various problems. In conclusion, a strong engagement of families, social services, and researchers is needed to put an end to the siege of this "empty fortress," so that autistic patients are no longer trapped inside themselves, which is all-pervasive but all-exclusive.

Chapter 3

Assistive Technology

3.1 Introduction

Technology-Related Assistance for Individuals with the Disabilities Act of 1988 defines assistive technologies as the range of products that are used to increase, maintain, or improve the functional capabilities of individuals with disabilities [24]. These technologies can be divided into two main categories: therapeutic and prosthetic. Therapeutic tools provide assistance in the form of teaching or therapy, with the aim of helping an individual deal with a disability or particular deficiency. Prosthetic assistive technologies, on the other hand, supplement or replace the natural capacities in which a person may have a disability. Prosthetic technologies provide people with various disabilities access to a wider range of resources and functions as well as help them integrate better in the community [24]. Many factors affect the choice of whether therapeutic or prosthetic technologies are considered for a particular disability. This depends

on the type of disability, its severity, a person's specific capabilities, and the affordability and social acceptability of the alternative technologies. It is also entirely possible that the two approaches be used to complement each other. Research suggests that people with autism generally feel comfortable in predictable environments, and more specifically, enjoy interacting with computers, e.g. (Colby and Smith [25], Powell [26], Moor [27]). One possible explanation has been put forward by Murray [28] who noted that the attention of people with autism tends to be fixed on isolated objects apart from the surrounding area. She argued that computers are the ideal resource to break into this world because they tap into the individual's attention tunnel which is focused on the screen and thus external events can be more easily ignored.

Many children with ASD have demonstrated that they can learn through these technologies [29]. This is shown by the evidence that some people with autism can tolerate higher sensory input via a computer than they can tolerate elsewhere. Moreover, computers offer a context-free environment in which many people with autism feel comfortable. Finally, computers are predictable and, are therefore, controllable. They enable errors to be made safely, encouraging children to use nonverbal or verbal expression. Therapeutic tools for autism include educational software, virtual environments, and the use of robotics.

3.2 Educational Software

Arguably, the greatest benefit of computers to individuals with ASD is that they offer a radically different mode of social engagement. The way people communicate is already changing, and this is potentially very beneficial to individuals with autism. Computer-mediated communication through e-mail, chat rooms and virtual reality environments may allow individuals with ASD to overcome some of their social problems. The aim of educational software is to introduce a interface between the patient and social interlocutor simplifying the communication. Moreover, it can be used to teach children the basic emotions.

3.2.1 Fun with Feelings

Fun with Feelings [30] (Figure 1) is a commercial software product that aims to teach children about the basic emotions (happy, sad, angry, afraid, surprised, and disgusted). Using a series of games, the children are taught the facial expressions and sounds associated with each of the basic emotions, and the situational factors that give rise to them. Each emotion is decomposed into several small pieces that are built in stages into an image of that emotion for the child. While general reviews of the software indicate popularity among autistic children, to the best of our knowledge, there are no scientific studies that evaluate the effective improvement in reading emotions.

Figure 3.1: Fun with feelings.

3.2.2 Bubble Dialogue

The Bubble Dialogue [31] application is a role playing computer program that allows users to reflect on speech as Dialogue and also alerts them to thought content as something distinct from speech. The program regulates turn-taking and serves as an interface between the two users. In this way Bubble Dialogue might allow the meeting of the minds to occur as an explicit, engaging and experiential process. Jones and Selby (1997) [32] used Bubble Dialogue to explore communication in children with emotional and behavioural difficulties. They suggest that Bubble Dialogue can help children to communicate and express their feelings and views when they find it difficult to communicate directly. This virtue of Bubble Dialogue might also apply to individuals with Asperger's syndrome. Jones and Selby also state that the role-playing element provides an emotional 'distance' which allows specific issues in the child's life to be raised without direct reference. Hence, participants could play characters

without having to closely identify with them.

3.2.3 Mental state simulator

Mental state simulator [33] is a software that uses formal AI reasoning to teach children diagnosed with High Functioning Autism about mental state concepts and representations. A child is encouraged to understand the initial mental state that includes the initial emotions of agents. Then the child is instructed to compose a story of the multi-agent behavior and compare it with what the software generates. In doing so, the trainees actively use a variety of mental entities, including basic, derived ones and emotions. Each column of the UI represents a mental state of an agent (in natural language on the top row, in the formal language - on the bottom row). Mental states can be modified either in the text area or using the row of combination boxes for each predicate name and its topics. Simulator builds the resultant scenario, and the trainee can analyze it and compare it with his own vision of the possible scenario. The test on two children show an overall improvement in both children with reasoning about deception, pretend play, and the expressions of surprise, sorrow and happiness. They were also able to understand jokes and to a lesser extent infer the intentions of others. It is worth pointing out that the two children were also attending other therapeutic sessions along with using the mental state simulator. It is therefore unclear whether the marked improvement can be entirely attributed to using the software. In addition, the study of two cases is a rather limited test-set, making it hard to draw conclusions about its success in helping these children learn about emotions beyond the

hypothetical scenarios presented in the program.

3.2.4 Mind-reading

Mind Reading is an interactive systematic guide to emotions, for its effectiveness in teaching adults with Asperger syndrome (AS) and high-functioning autism (HFA) in order to recognize complex emotions in faces and voices [34] [35]. It is based on a taxonomic system of over 400 emotions and mental states, grouped into 24 emotion groups, and six developmental levels (from age 4 to adult). The "Mind-reading" was developed by a team of psychologists led by Professor Simon Baron-Cohen at the Autism Research Centre at the University of Cambridge, working closely with a London multimedia production company. Each emotion is captured through six audio and video clips and is also explained through six stories to give a flavor of the kinds of contexts that give rise to that emotion. The game zone encourages informal learning about emotions in a less structured setting. The games tackle a number of contexts for reading emotions including a school, an office and a marketplace.
Results of the experiments revealed that the use of Mind Reading led adults with autism spectrum conditions to improve significantly in their emotion recognition skills. This improvement was achieved over a relatively short period of time on a variety of complex and socially important emotions and mental states, in both faces and voices. The improvement was made also independently from additional tutor and group support.

Figure 3.2: Screenshots from Mind Reading: The Interactive Guide to Emotions [34].

3.3 Virtual Environments

A Virtual Environment (VE) is a computer-generated, 3D environment that aims to surround, or immerse the user so that he or she becomes part of the experience in a simulated environment [36], potentially allowing the transfer of skills from the virtual to the real world [24]. Virtual environments can be widely used as an educational medium. This is a particular issue when the intended users are people with disabilities or have severe learning difficulties (mental handicap or mental retardation), for whom the generalization of skills from training situations to the real world is notoriously difficult. With a combination of a careful design and programmatic use directed at clear and obtainable learning objectives, they can be made to mimic closely the salient or critical features of the real world, making it more likely that skills will be transferred successfully. Where risk is a factor they permit the user to learn by making mistakes without suffering the real consequences of their errors. They may also have a particular pertinence in the training of people with learning disabilities because they promote self-directed activity, decrease reliance on language or other abstract symbol systems, and minimize the effects of physical impairments [36].
. Many virtual games have been proposed by Strickland [37], which teach children diagnosed with autism about a range of social situations. He claims that virtual reality offers the potential to regulate an artificial computer environment to better match the expectations and needs of individuals effected by autism and attention disorders. Strickland [37]verified the acceptability of the VR equipment and the verbal response of the games. His results indicated an encouraging

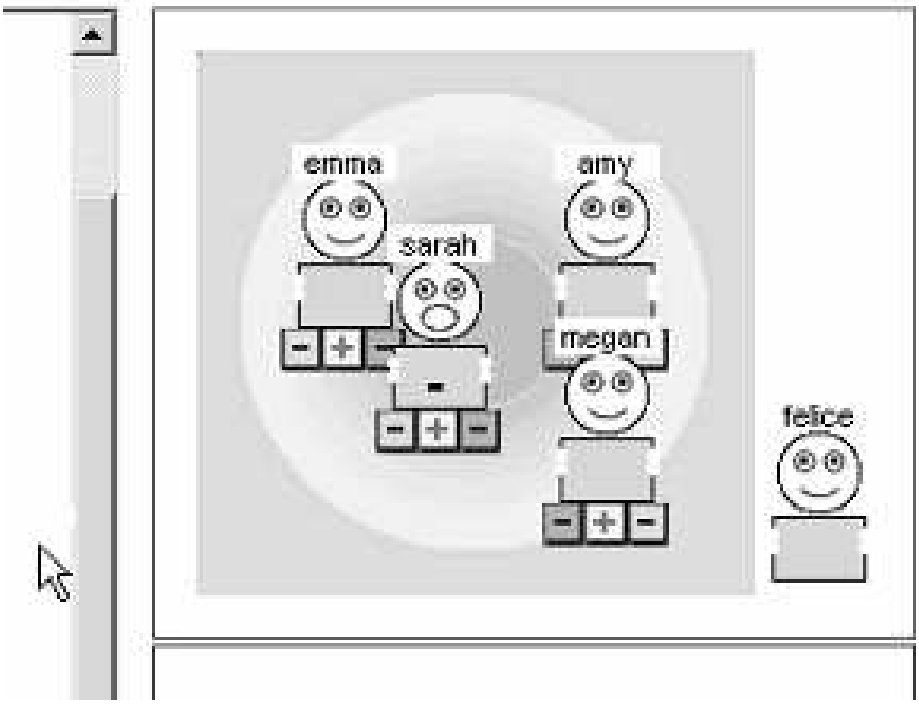

Figure 3.3: Therapist Client User Interface, Position Tool [38].

adaptation to the technology by two children who weared a virtual helmet. Cheng et al. created Kidtalk [38], an online environment used to treat children with Asperger's Syndrome/High Functioning Autism. The environment is presented as an online chatting environment, where children work through common social situations, such as going to the movies or going to a party, by chatting online. Moreover, a richer setup for children to explore different social scenarios compared with educational or training software is provided. However, similar to educational software, more studies are needed

to investigate whether learning experiences through these virtual environments can be generalized to represent to other real-world environments [24].

3.4 Robotics

In addition to computer-based technologies, several therapeutic tools are based on the use of robots. In fact, in recent years, robots have been proposed for the treatment and rehabilitation of people with cognitive disorders. The driving force was the emergence of epigenetic robotics [39], a fusion between psychology and robotics . A robotic cognitive system grows as a result of interaction with the social and physical environment [40]. Recent research has shown that certain subjects with cognitive deficits perceive and treat robots not as machines, but as their artificial partners [20] [21]. Moreover, social and emotional reactions may be transmitted to individuals with ASD via a three-dimensional display based on social or cybot robots [22] [23]. This opens a new issue: What aspects in the robot's appearance can facilitate interactions that might encourage basic social interaction skills? This aspect also leads to the issue of believability, an argument which is largely explored in the Japanese robotic community under the umbrella of "the uncanny valley" paradigm [41]. Based on these observations, several robotic artifacts have been used to engage proactive interactive responses in children with ASD. The following sections report the different approaches and results.

3.4.1 Robots in Aurora Project

The AuRoRA Project studies if and how robots can become a "toy" that might serve an educational or therapeutic role for children with autism. Their main aim is to engage children with autism in coordinated and synchronized interactions with the environment, thus helping them to develop and increase their communication and social interaction skills. Important social skills that they aim to facilitate are turn-taking and imitation, in addition to general communication and interaction skills that are required in human-human contact. Humans are the best models for human social behavior, but their social behavior is very subtle, elaborate, and widely unpredictable [22]. Many children with autism are, however, interested in playing with mechanical toys or computers. Different robots have been developed during the project illustrating the ability of the robots to provide a focus on attention, and shared attention in trials with pairs of children with autism.
KASPAR [42] is a child-sized robot which acts as a platform for HRI studies, using mainly bodily expressions (movements of the head, hand, arms), facial expressions, and gestures to interact with a human (Fig. 3.4). The robot has a static body (torso, legs and hands do not move and were taken from a child-sized commercially available mannequin doll) with an 8-DOF (degree of freedom) head and two, 3-DOF arms. The important features of Kaspar's head are its minimal design, the inclusion of eyelids, and the aesthetic consistency of the face [43] [44]. The overall design rationale of Kaspar's head and face aims to approximate some important features of the appearance and movements of a human without trying to create an

Figure 3.4: Some of KASPAR's facial expressions, and expressive gestures

ultra-realistic appearance, i.e. not trying to imitate every detail of a human face (see Fig. 3.4).

The children show a level of direct, physical engagement with KASPAR, but also they appear to generalize this behavior at least to those who were present during this study. Thus, the children appear to use touching and gazing at KASPAR prior to touching and gazing at the others present (Fig. 3.5).

Furthermore, children appear to show the researchers present during the experiment some awareness perceptions of Kaspar, turning to gaze at them following some potentially relevant action by KASPAR that they are treated as having perceived. This robot is an example of how (relatively) simple, low-budget, minimally expressive robots can facilitate interactive 'social' games that benefit the children and the wider social environment. This clearly emphasizes a 'bottom-up' approach where first the specific needs of the children and the application environment were considered, which then led to the use of an appropriate robot that matches the needs in this application domain. Thus, the behavioural and structural complexity

Figure 3.5: A child engages in a tactile exploration of KASPAR

of the robot is matched with the needs in the 'niche' application. The Robota project [45] [23] is another among several approaches, including the Aurora project, that investigates the potential of a robotic platform as part of educative therapy for children with autism. Robota complements other approaches by addressing the role that human features and imitation might have in shaping children's interaction with others.

Robota systematically assesses the child's social skills through a rich spectrum of interactions: speaking, music and movements. Robota, a doll robot (3.6), is 45 cm tall and has many motors in order to move the head and the legs. Plastic arms are attached to the body. The doll may show many facial features such as human skin. The robot can react to touch through the use of a potenziometer in order to detect the passive movements of the arms and head. Using movement tracking, Robota can mimic the end user extending the interaction in imitational games. The learning algorithm provides the robot with the ability to communicate using vocal commands. The experiments have showed that after the familiarity stage, the children start to interact with the robot sharing the game experience. With Robota, the reported studies led to replace the early "invasive" infrared set-up by a "noninvasive," vision-based mechanism for the imitation game. The vision-based interface was better tolerated by the children. Moreover, evidence suggesting that facial features have an impact on the children's willingness to interact led to the development of a prototype with mobile eyes that will be used subsequently to test the role that eye-gaze may have on the children.

Figure 3.6: Robota robot. A camera mounted on the PocketPC of Robota enables tracking of the hands and nose of the user through optical flow and recognition of skin colour. Robota can then mirror the movements of the arm and head of the user [45]

3.4.2 Storytelling robots

The storytelling robots have been developed for use with children undergoing rehabilitation by Plaisant [46]. This storytelling robot was originally designed for all children, without consideration for special challenges. Children using the storytelling software could write stories and select movements or "emotions" from a pre-programmed list (e.g. happy, sad, surprised). The story could then be "played" by the remote controlled robot, which told the story (playing recorded or synthesized speech) and acted out the movements at the appropriate time in the story. Afterwards, the research scope was extended the tool for the children's rehabilitation. The system utilizes a variety of sensors connected to detect input (or "gestures") from the user (3.7).

Software was then used to assess and process these inputs and to map the inputs to appropriate outputs. From the design sessions with children, it emerged that the storytelling robot was a very powerful tool to engage children in discussing storytelling, proving that PETS could be used as a rehabilitation tool for children with developmental disabilities. Storytelling appeared to offer a solid motivating direction as a typical home activity (with relatives, siblings or friends). Moreover, play can give children with disabilities a sense of competence and control over environmental circumstances that help them learn new skills.

3.4.3 Infanoid and Keepon

Kozima and Yano [39] investigated the possibilities of using humanoid robots in therapy. They built a child-like humanoid, In-

Figure 3.7: Sensors mounted in arm bands and in a hat allow a user to directly control the robot [46]

fanoid, and a small creature-like robot, Keepon, which are capable of primordial embodied communication with humans, especially children and babies, respectively. The approach implements psychological models of social development in the robots in order to observe how human children interact with the robots through various social actions, such as pointing at and showing objects. Through observations, they upgrade the robot's structure and function, thus forming an elaboration cycle. This mode of operation fits with epigenetic robotics where complex cognitive and perceptual structures emerge as a result of the interaction of an embodied system with a physical and social environment [47] [48].

Infanoid, shown in Figure 3.8, is an upper-torso humanoid robot which is as big as a 3- to 4-year-old human child [47] [49]. It has 29 actuators (mostly DC motors with encoders and torque sensing devices) and a number of sensors arranged in the relatively small body. It has two hands capable of pointing, grasping, and a variety of hand gestures; it also has lips and eyebrows to produce various facial expressions, like surprise and anger. The head of Infanoid has two eyes, each of which contains two different color CCD cameras for peripheral and foveal views; the eyes can perform saccadic eye movements and smooth pursuit of a visual target. From the microphones at the ears, Infanoid hears human voices and analyzes the sound into a sequence of phonemes. By feeding the phoneme string with the extracted pattern of fundamental frequency, Infanoid performs vocal imitations while sharing attention with the interaction. For psychological experiments with younger children, they built another robot, Keepon [50]. Compared with Infanoid, which has a human-like sophisticated structure and function, Keepon is a small

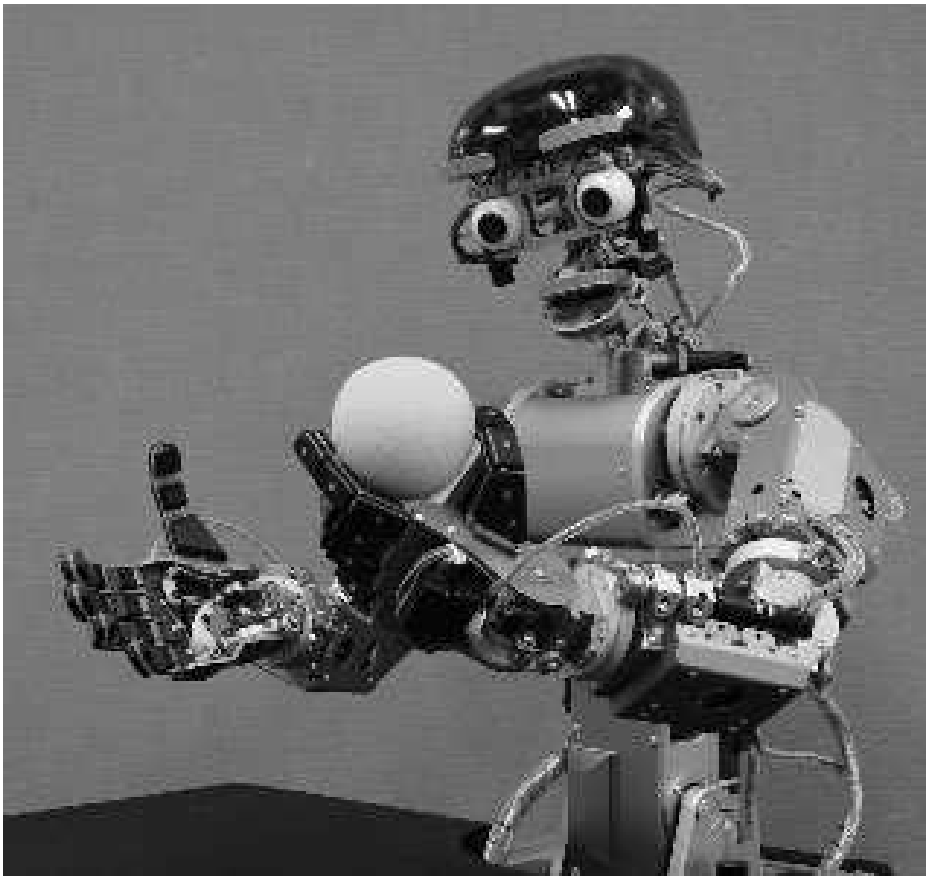

Figure 3.8: Infanoid robot [48].

(12cm in height, 8cm in diameter), soft (made of silicone rubber), creature-like robot, as shown in Figure 3.9.

Keepon is able to explicate the attentive state moving the face towards an object. Moreover, it is able to express the emotive state, like pleasure and excitation, by movements of its body (rocking). The robot has been designed to achieve very simple emotional and attentive exchanges with a human agent. Therefore, it attracts the attention of the children generating spontaneous reactions. The robot head has two eyes where a CCD camera is installed with a very large visual range. In the body, motors and the control system are mounted. The 4-DOF of the body provides the robot with the capability to execute different actions:

- Attentive action in order to direct the body to a object or to enable the visual contact;
- Emotional action: Maintaining its attention, Keepon shakes the body to mimic pleasure and excitation.

The restricted number of actions provides the user with a very simple communication, so even a child is able to recognize the perception of Keepon. The robot is designed to work in manual operation or automatic operation. In the latter, the intelligence software enables the robot to track the position of the human visage or of the object with a particular color. However, the robot usually operates in a manual configuration where an operator moves the robot, controls the voice, and selects a facial expression according to the situation. The results have shown that the children interacted with curiosity and assurance sharing their surprise and pleasure with educators,

Figure 3.9: Keepon robot [50].

Figure 3.10: Paro robot

teachers and coetaneous. The storage data are used by the therapist to improve the therapeutic services. The child-like humanoid, Infanoid, and the creature-like robot, Keepon, are being used as research platforms on which test models of social intelligence and its development are implemented.

3.4.4 Paro

Paro is an assistive tool in rehabilitation and robot assisted activity introduced by Shibata et al.[51] [52]. Paro was proposed as a tool that could benefit elderly people, hospitalized children, as well as children with autism. The robot was inspired by arctic seal pup (Fig.3.10).

This choice was based on previous efforts to replicate the behaivour of a cat and dog, which did not support good interactive dynamics [51]. Paro has the following senses:

- Light sensors;
- Directional acoustic sensors;
- Position sensors;
- Haptic sensors.

However, in this work, very little was documented about the particular history of the children and the specific nature of the therapeutic effects that can be linked to the robot e.g. what types of robotic behavior were beneficial to the child, and what types of therapeutically relevant behaviors were targeted.

3.5 Wearable Devices

There has been an increasing surge of interest in developing wearable devices with the goal of applying them in applications of augmented reality and perceptual interfaces, especially vision-based interfaces. In augmented reality, computer-generated images are superimposed over a user's view of the physical world. Perceptual interfaces are technologies that have perceptive capabilities built into them, vision being one of them. Vision-based interfaces perform a range of tasks including people recognition, identity recognition, gesture recognition, facial expression for emotion recognition, and object detection and video surveillance. In assistive technology, wearable devices have

been proposed to assist those with cognitive impairments, such as traumatic brain injury and dementia. For example, COACH (Cognitive Orthosis for Assisting activities in Home) [53] aims to develop automated computer vision systems to assist patients. Such systems would be able to non-invasively monitor the patient, stepping in to provide help in the form of verbal or visual prompts when necessary. Recently, a particular interest has been addressed to develop wearable eye tracking system to observe behavioral characteristics of individuals with ASD. In fact, there is a growing body of research that makes use of eye-tracking technology to study attention disorders and visual processing in ASD. Atypical gaze patterns were already described for individuals with ASD when presented with social scenes and faces [54] [55]. For instance, Klin et al.[56] pointed out that, in social environments, individuals with autism show reduced eye-region fixation time in favour of an increased focus on mouths and objects. Reduced attention to the face but not to the actions of a demonstrator to be imitated was observed by Vivanti et al. [55] in a group of children with autism. Gaze tracking could be thus a critical and useful indicator of a subject's interest and emotional involvement during a therapeutic session.
Based on these studies, Billard et. al introduced a low intrusive wearable eye tracking named Wearcam [57]. The WearCam is one of the therapeutic devices developed within the TACT (Thought in Action) research Project, financed by the European Union's NEST'Adventure Program. This project aims at developing non-obtrusive monitoring devices with toys appearance for the study on young children behavioral characteristics like: movement, attention, voice, grasping force, etc. The WearCam is a wearable wireless camera located on the

forehead of the child, (Fig.3.11). It collects video recordings during therapeutic sessions as well as free play sessions in which the child takes part.

Wearcam has been useful to investigate whether possible differences in visual attention when observing an action to be imitated may contribute to imitative difficulties in autism in both non-meaningful gestures and meaningful actions on objects.

Figure 3.11: A normally developing child wearing a prototype of the WearCam

Chapter 4

FACE, an embodied interactive social interface

Human-like robots that embody emotional state expressions, empathy and non-verbal communication have been proposed for autism therapy. They can be thought of as a sort of robot-based affective computing. The University of Pisa group has pioneered this approach through the use of a lifelike android FACE (Facial Automation for Conveying Emotions), which displays emotional information through non-verbal communication [58]. The hypothesis is that adaptive therapy using a robot endowed with the ability to sense, adapt and respond to a patient's postulated emotional and mental states will enable autistic subjects to learn empathy and gradually enhance their social competence. In particular, the therapy could help autistic subjects to interpret emotional states of an interlocutor, through familiarity and contextual information presented in a stepwise and controlled manner. The FACE robot is capable, in its

Figure 4.1: Version II: FACE robot produced by Academy of Fine Arts of Carrara.

present embodiment, of mimicking a limited set of facial expressions, which are more easily accepted by autistic patients because of their simple and stereotypical nature. In this scenario, after an initial training period in which the patient familiarizes himself with the robot, different emotional "mood states" are created in the android modulating its behaviour, influencing the expression selection and intensity as a function of the subjects postulated emotional states [59]. As the process of imitation plays a crucial role in distinguishing between actions arising from within and actions induced by others, it is an important part of the therapeutic process. It paves the way of the comprehension of others' intentions by establishing a reciprocal, non-verbal communication process in which the roles of imitator and reference are continuously exchanged. Moreover, imitation plays a fundamental role for the emergence of proprioception, of the perception of the external world and of the ability to act out our own actions as well as those of others. True adaptive therapy entails control and modulation of the expressions and non-verbal actions of FACE by feedback from the emotional reactions from the subjects themselves, postulated by capturing and analysing their physiological and behavioral correlates through sensorised wearables. Using this closed looped approach the path towards learning and processing emotional understanding is orchestrated by a therapist, but finely tuned by the ASD subject's responses. In adaptive therapy the treatment is automatically synchronized to the subject's emotional needs and learning pace without being overly intrusive or insistent. The range of autism spectrum disorders as well as other developmental deficits is huge, and it is likely that individuals respond to behavioral and psychological therapy differently, much as they respond to drugs. Therefore

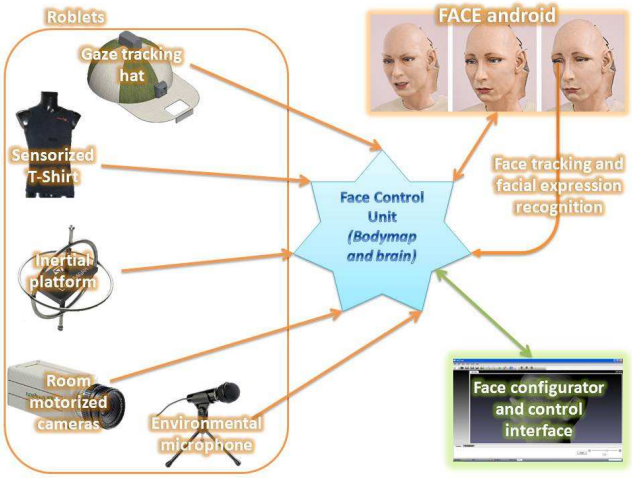

Figure 4.2: Connection scheme of the FACET platform. In orange there are the different roblets that communicate with the FACE control unit where the brain and bodymap are hosted in blue. The unit can also host the evolvable hardware as a roblet. Supervision and therapist control are allowed through the FACE configurator roblet in green [59]

a structured therapeutic environment is used in which the subject's behavior and responses are monitored using a multiplicity of sensors and then processed and fed back to the android to modulate and modify its expressions (Fig.6.1).

The integrated sensing, monitoring, processing and emotionally responsive android-based therapeutic platform is termed FACET (FACE Therapy). Physiological signals (HR, HRV, RR, EMG and EDA) and eye gaze are acquired using a comfortable, unobtrusive

Figure 4.3: Version III: FACE robot produced by D. Hanson

sensorised shirt and a cap containing integrated cameras and mini gyroscopes.

4.1 Face Robot Hardware

FACE is an android used as an emotion conveying system. It consists of a female face made of FlubberTM, a skin-like silicone based rubber patented by Hanson Robotics (Fig.4.3).

Android faces produced by D. Hanson have been used in other robots, with their own software architectures, such as the Ibn Sina Robot [60], Javier Movellan's robot at UCSD [61], and the INDIGO project [62]. FACE is actuated by 32 servo-motors that move the

artificial skin through cables inserted in the face according to the human anatomy. The motor cables act as tendons moving FACE's skin and allowing human facial expressions to be re-created. FACE servo-motors are all integrated in the android skull except for the 5 neck servos that allow pitch, roll and yaw movement of the head. The android has a CCD camera in the right eye used for face tracking of the subject.

4.2 The cognitive architecture

The multiplicity of sensors introduced the need to have an Artificial Decision Unit (ADU) able to classify the visual input and physiological signals as well as to actuate the servo-motors embedded into the android (Fig.4.4). Therefore a cognitive architecture reproducing the functionalities of brains such as the Artificial neural networks (ANNs), which are computational models implemented in software or custom-made hardware devices that attempt to capture the behavioral and adaptive features of biological nervous systems [63], appear to be the best solution. Moreover, the real-time interaction with the patient requires that the system have the following:

- Computational Complexity
- Speed performance
- Small size
- Low power consumption

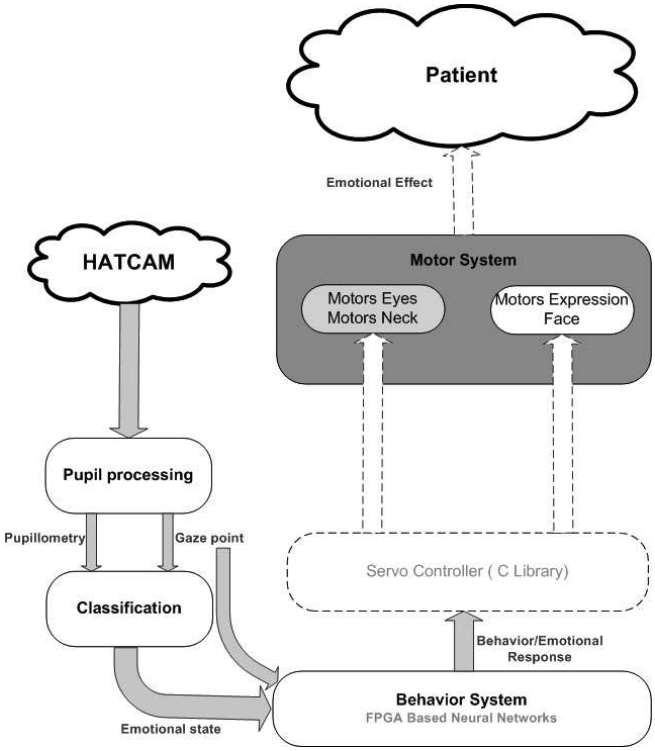

Figure 4.4: Control system considered in this work. The image captured by the eye tracking system named HATCAM is processed in order to obtain information about the pupil size and gaze point of the patient. This information is an input of the Behavior System.

The best architecture that provides a good compromise between all requirements is the implementation of the ANN on a Field Programmable Gate Array (FPGA). In fact, the FPGA provides a good trade-off between speed performance, flexibility and power consumption. Moreover, ANN implemented in hardware have an important advantage over computer simulated ANNs because they fully exploit the parallel operation of the neurons, thereby achieving a very high speed of information processing [64] [65]. The following sections will provide an in depth description about the implementation of ANN starting from the implementation of the single neuron to the architecture topology of the neural network.

4.3 Neural Networks in Hardware

Processing architectures are often performed by software models inspired by biology, such as artificial neural networks (ANN). These architectures must perform high-efficiency complex tasks such as image processing, data filtering, feature selection, multimodal perception, control systems etc. In addition, these tasks may introduce module interconnection and cooperation issues, as well as large execution time, negatively affecting real-time applications. Nevertheless, ANN hardware implementations may take advantage of their inherent parallelism allowing faster responses than software implementation [64] [65]. Moreover,the recent advent of dense, high-speed FPGAs implies that hardware neurocomputers are more feasible thanks also to the high degree of reconfigurability of the devices [66].
In the hardware implementation of ANNs, three main design met-

rics are relevant: accuracy, gate complexity and processing speed [64] [67]. The design of hardware implementations is governed by balancing these criteria. At instance, the requirement of very high precision cannot be matched with fast and compact hardware implementations [68] . Nevertheless, a low precision influences the convergence properties of the network. It is generally accepted, indeed, that, as the resolution and precision of the inputs, weights and the activation function are reduced, the capability of the ANN of acting as a universal approximator is reduced as well [69] [70]. However, there are currently many studies (e.g. [70] [71] [72]) that have explored this problem.
The most complex task to implement the hardware ANNs is the non-linear activation function. Common examples of activation function include hard-limiter, saturated linear, hyperbolic tangent function and sigmoid function. The following subsections will face the digital implementation of sigmoid and hyperbolic tangent functions. A new mathematical approach inspired to "area method" originally proposed by Francesco Paolo Cantelli in the 1905 [73] is introduced. Moreover, an approximation of the hyperbolic tangent and its derivative is extrapolated by the mathematical relationship between the sigmoid and the hyperbolic tangent. So, the sigmoid is used as generator of hyperbolic tangent with only subtractions and shifts. The hardware implementation of the two functions is realized.

4.3.1 State of the Arts

The most common non linear activation functions, which are used in the artificial neural networks, are the sigmoid function and the

hyperbolic tangent. Both the functions are included in the logistic functions introduced in the economic and demographic studies by Verhulst [74]. These functions are mainly used in statistics ,biomathematics, physics, engineering, economic science etc. The general equation is as follows:

$$y = \frac{a}{1 + e^{-bx+c}} + d \tag{4.1}$$

where a,b,c and d are constants.
The sigmoid function is a particular case of equation (4.1) where we put a=1, b=1, c=0 and, d=0. The equation thus becoming:

$$y = S(x) = \frac{1}{1 + e^{-x}} \tag{4.2}$$

On the other hand, when a=2, b=2, c=0 and d=-1, the equation represents the hyperbolic tangent:

$$y = T(x) = \frac{2}{1 + e^{-2x}} - 1 = \frac{e^{2x} - 1}{e^{2x} + 1} \tag{4.3}$$

The direct hardware implementation of (4.2) and (4.3) is not friendly because it requires a lot of logic elements. For this purpose a large number of approximation techniques have been developed. The performance of these techniques is evaluated by absolute and relative errors. Moreover, as the number of elementary logic components and the memory load are reduced, performance increases. This mean that many of the techniques that have been developed in numerical analysis e.g. summing of truncated series expansions, are expensive in terms of computation time for hardware implementation [75] [76] [77].

The common approximation methods are based on CORDIC (COordinate Rotation DIgital Computer) algorithm, linear polynomial approximations or high-order polynomial approximations, approximation by the ratio of two polynomials, approximation by lookup table, hybrid techniques.
The CORDIC algorithm, usually applied to calculate the approximation of the hyperbolic tangent, was introduced in 1959 by J. Volder [78] for the computation of the trigonometric function by means of additions and shifts. The follow extensions of the CORDIC in linear and hyperbolic coordinates, based on formulas similar to those in circular coordinates, suggested a unification introduced by Walther [79].
Nevertheless, the CORDIC algorithm does not assure the convergence of the hyperbolic function unless repeating some iterations and for limited input domain. This problem has been solved by A.Y.Hu [80] who introduced an iteration with negative index but the hardware implementation is quite complex.
The main technique used to approximate the activation function is the piecewise linear approximation (PWL). The input range of the function is divided into equal parts and, in each subinterval, the function is approximated by a straight line. The method has low computational load with low average and max absolute error. This technique has been used with different levels of approximation in [75] [81] [82] [83] [84] [85] [77] for the sigmoid function and in [76] [86] for the hyperbolic tangent.
Methods with high-order polynomials, e.g. second order, are not convenient due to the high number of multiplications. Zhang [83] have presented a technique requiring a multiplication, two shifts and two

XOR.
The approximation based on the ratio of two polynomials are not practically convenient for the presence of the division, while the function tabulation is costly in term of memory. A different addressing was proposed in [87]. In detail Lebouf proposed a RA-LUT method to approximate the hyperbolic tangent reducing half comparator and therefore area at the expense of accuracy. Lin introduced the possibility to approximate the first-order derivative of the original function and then to obtain the original function by integrating the first-order derivative approximation based on Isosceles Triangular Approximation (ITA) or PWL [86]. A method to approximate the sigmoid function, based on recursive interpolation is the CRI (Centred Recursive Interpolation), which has been presented by Basterextea [67].
A direct bit level mapping method has been proposed by Tommiska [64].

4.3.2 A new approximation method

The method used to approximate the sigmoid function, the hyperbolic tangent and their derivatives, originally introduced F. P. Cantelli [73], is based on a piecewise linear approximation. Cantelli stated that one of the most important problem of the observation science was to determine coefficients of the curve equation able to approximate a set of measures or observations, when the analytic form of this equation is known. Such a problem can be solved by the least-square method which, anyhow, implies very complex equation

systems. To overcome these difficulties, Cantelli proposed a new method, named area method, which shows good adaptation with simple computations:
In order to adapt a curve, opportunely chosen:

$$y = \varphi(x, c_1, c_2..., c_n). \tag{4.4}$$

to a set of observations, you have to divide the domain of the x-axis in n equal intervals determining the constants able to gain, in each of these intervals, the same area between the theoretical curve and the evaluated one by means of the observations.
The goodness of the approximation is related to the method used to calculate the areas obtained from the observations (Trapezoid, Simpson, etc.), to the approximation curve and to the divisions of the x-axis domain.
To make the method systematic, Cantelli divided the range of observations into equal intervals. Nevertheless, the orderliness of the method can be preserved with different choices as hypothesized by G.Castelnuovo [88] and as it will result from our hypothesis in the follow.
Modifying the area method, it is possible to approximate a known function, integrable in its range of definition. So we enunciate a modified method as follows:

To adapt a curve, opportunely chosen

$$y = \varphi(x, c_1, c_2..., c_n) \tag{4.5}$$

to an integrable function, you have to divide its domain into adequately individuated n intervals. You have to determine the constants so that, in each of these intervals, the partial areas subtended by the theoretical curve and those subtended by the approximation function, are equal

We will apply the modified method in order to achieve a PWL of the activation functions and their derivatives. The basic approach implies several steps.
Let f(x) a real function of equation (4.6), defined and integrable in the real range [A,B] ($A < B$).

$$y = f(x) \tag{4.6}$$

Subdividing the range [A,B] into equal parts n of amplitude $h = \frac{B-A}{n}$, let us approximate the (4.6) in each subinterval [a,b] by a linear function:

$$\varphi(x, c_1, c_2) = c_1 x + c_2. \tag{4.7}$$

Since (4.7) depends on two constants (c_1 and c_2), in order to apply the area method, each range [a,b] has to be divided in two parts: [a,c] and [c,b]. At the first application of our algorithm, the variable c can be selected in the]a,b[range without constraints. Tracing the perpendicular to the x-axis to point c, two trapezoids are obtained for the f(x) function as well as for the straight line of (4.7) (Figure 4.5). To achieve an ideal approximation must be:

$$\int_a^c f(x)dx = \int_a^c (c_1 x + c_2)dx \tag{4.8}$$

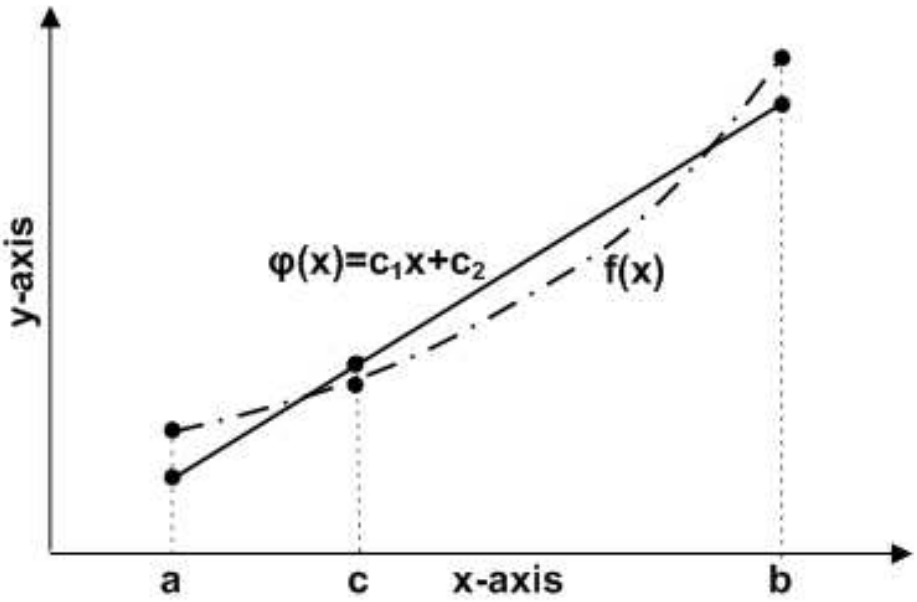

Figure 4.5: Approximation with area method

$$\int_c^b f(x)dx = \int_c^b (c_1 x + c_2)dx \tag{4.9}$$

The equations (4.8) and (4.9) represent two equations with unknown quantity c_1, c_2 that integrated and solved, allow us to univocally determine the straight line approximating f(x) in the range [a,b].

In order to evaluate the performance of the method and compare it with the state of the art, the absolute and relative errors are calculated for 10^6 input data uniformly spaced in the domain [A, B] as done in [83]. For a function f(x), approximated by the straight line

$\varphi(x)$ in the domain [A,B] the average and maximum absolute errors are defined as:

$$E_{ave} = \sum_{i=0}^{10^6-1} \frac{|f(x_i) - \varphi(x_i)|}{10^6}. \tag{4.10}$$

$$E_{max} = max|f(x_i) - \varphi(x_i)|. \tag{4.11}$$

Similarly, we define the average and maximum relative errors as:

$$E_{averelative} = \sum_{i=0}^{10^6-1} \frac{|f(x_i) - \varphi(x_i)|/f(x_i)}{10^6}. \tag{4.12}$$

$$E_{maxrelative} = max\frac{|f(x_i) - \varphi(x_i)|}{f(x_i)}. \tag{4.13}$$

Referring to the orderliness of the method, we have verified that the choice of the point c as middle point of the range [a,b] not always brings to a good approximation. Therefore, according to Castelnuovo [88], who introduced the possibility to divide the range [a,b] into non equal parts, we have modified the area method identifying the c point inside the subinterval [a,b] in order to obtain the minimum of the maximum relative error (or the minimum of the maximum absolute error) without losing the generalization of the method .

To obtain the variable c, the range]a,b[is subdivided in 10^3 equal parts and for each point $c^i = a + \frac{i \cdot h}{10^3}$ $(i = 1, 2, 10^3 - 1)$ we solve the (4.8) and the (4.9) thus calculating c_1^i and c_2^i, and hence the straight line $\varphi = c_1^i x + c_2^i$ is determined. By the (4.13) and for every point inside the range [a,b] the max relative error (or absolute error) is

calculated. When the cycle ends the min max relative error (or min max absolute error) is chosen. The method, repeated for all division subintervals of range [A,B], provides a piecewise linear approximation of the f(x). In our applications, the first subinterval of division [A, A+(B-A)/n] is considered close while the others are considered open on the left]a,b], to preserve the univocally of the function.

4.3.3 Sigmoid approximation

The study of the sigmoid function (4.2) is typically limited to the range [-8, 8], where it is continue and integrable, but since the sigmoid is a symmetric curve with respect to the point having coordinates (0, 0.5), the study can be reduced to [-8, 0]. Dividing the range [-8, 0] into unitary subintervals, and applying the area method in each of them we have:

$$\int_a^c \frac{1}{1+e^{-x}}dx = \int_a^c (c_1x + c_2)dx. \tag{4.14}$$

$$\int_c^b \frac{1}{1+e^{-x}}dx = \int_c^b (c_1x + c_2)dx. \tag{4.15}$$

The unknown quantities c_1 and c_2 in (4.14) and (4.15) will be determined in order to obtain the straight line:

$$S_{approx}(x) = c_1x + c_2. \tag{4.16}$$

Integrating both members of the (4.14) and (4.15) we have the following system:

$$\begin{cases} (\frac{c^2}{2} - \frac{a^2}{2})c_1 + (c-a)c_2 = ln\frac{1+e^c}{1+e^a} \\ (\frac{b^2}{2} - \frac{c^2}{2})c_1 + (b-c)c_2 = ln\frac{1+e^b}{1+e^c} \end{cases}$$

Solving for c_1 and c_2:

$$c_1 = \frac{(b-c)ln\frac{1+e^c}{1+e^a} - (c-a)ln\frac{1+e^b}{1+e^c}}{0.5(a-b)(b-c)(c-a)}. \tag{4.17}$$

$$c_2 = \frac{(c^2-a^2)ln\frac{1+e^b}{1+e^c} - (b^2-c^2)ln\frac{1+e^c}{1+e^a}}{(a-b)(b-c)(c-a)}. \tag{4.18}$$

The straight line $\varphi(x) = c_1x + c_2$ is therefore univocally determined. Then, we choose the c point in order to obtain the minimum of the maximum relative error in every subinterval]a,b[as reported in the Section 4.3.2. The Fig.4.6 plots, at instance, the maximum of the relative error in the range]-8,-7[in function of c. The minimum is obtained for c=-7.678.

The results of the sigmoid function are obtained and shown in Table 4.1, the curve in Fig.4.7 and the relative error in Fig.4.8.

Table 4.1: Polynomial approximation of the sigmoid function and c subdivision points

Range	c_1	c_2	c	**Range**	c_1	c_2	c
[-8,-7]	0.00053	0.00456	-7.678	]0,1]	0.22830	0.50596	0.729
]-7,-6]	0.00144	0.01094	-6.678	]1,2]	0.14945	0.58948	1.683
]-6,-5]	0.00390	0.02571	-5.678	]2,3]	0.06749	0.75292	2.678
]-5,-4]	0.01046	0.05852	-4.678	]3,4]	0.02739	0.87367	3.678
]-4,-3]	0.02739	0.12632	-3.678	]4,5]	0.01046	0.94147	4.678
]-3,-2]	0.06749	0.24707	-2.678	]5,6]	0.00390	0.97428	5.678
]-2,-1]	0.14945	0.41051	-1.683	]6,7]	0.00144	0.98905	6.678
]-1,0]	0.22830	0.49403	-0.729	]7,8]	0.00053	0.99543	7.678

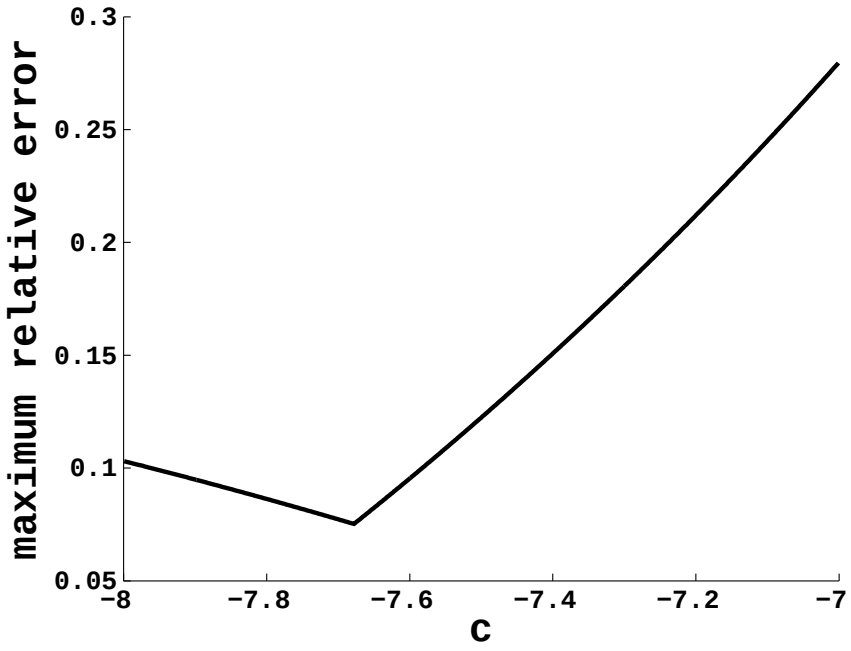

Figure 4.6: Maximum of the relative error in function of c

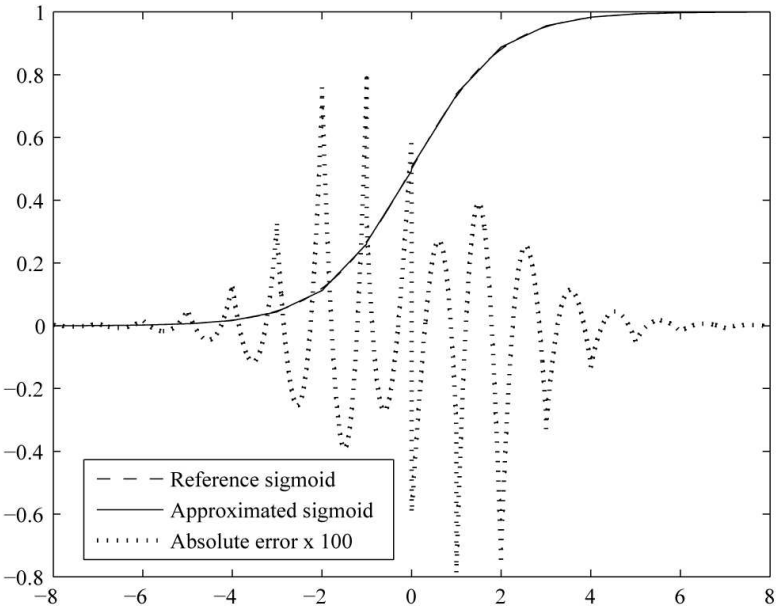

Figure 4.7: Approximated sigmoid function

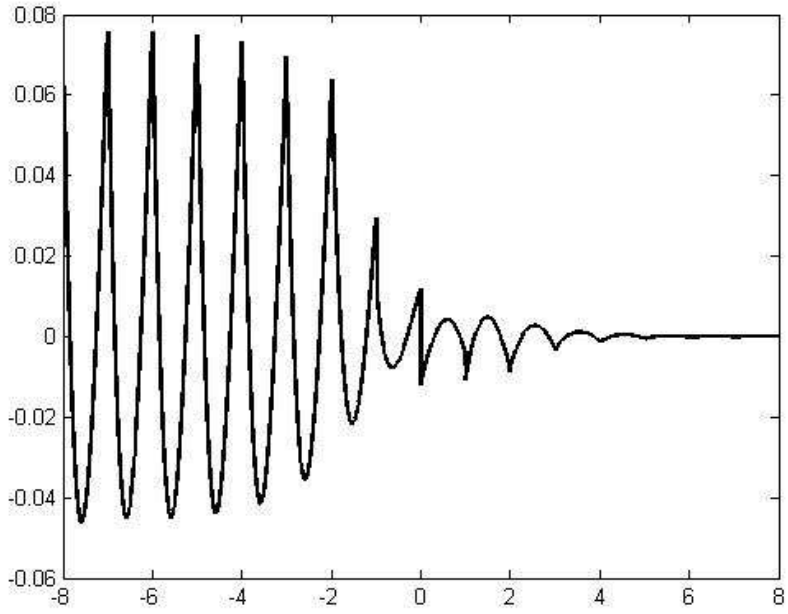

Figure 4.8: Relative error of the sigmoid function

The performance of the method was compared with the main approximation studies [64] [67] [75] [81] [83] [84] [77] (Table 4.2).

Table 4.2: Summary of absolute and relative errors in the published literature for the sigmoid

Method	**Input range**	**Absolute Error**		**Relative Error**	
		$E_{ave}\%$	$E_{max}\%$	$E_{ave}\%$	$E_{max}\%$
Area Method	**[-8,8]**	**0.1070**	**0.7880**	**1.349**	**7.503**
Bajger [77]	[-8,8]	0.1199	0.8616	1.547	6.147
Faiedh [84]	[-5,5]	0.20	1.11	-	-
Myers [75]	[-8,8[	2.47	4.90	-	-
Alippi [81]	[-8,8[	0.87	1.89	-	-
PLAN [82]	[-8,8[	0.59	1.89	-	-
CRI q=3 [67]	[-8,8[	0.85	2.06	-	-
Tommiska [64]	[-8,8[	0.17	0.39	-	-
Zhang [83]	]-4,4[	0.77	2.16	-	-

To implement gradient-descent-type learning algorithms, the same method has been applied to approximate the derivative of the sigmoid function:

$$\frac{dS(x)}{dx} = S(x)(1 - S(x)) = \frac{e^{-x}}{(1 + e^{-x})^2} \tag{4.19}$$

which is continue and integrable in [-8,8] and symmetric with respect to the y-axis, applying a further subdivision in the range [-1, 0] for a better approximation.
The detailed piecewise approximation results of the derivative function are shown in Table 4.3 and in Fig. 4.9.

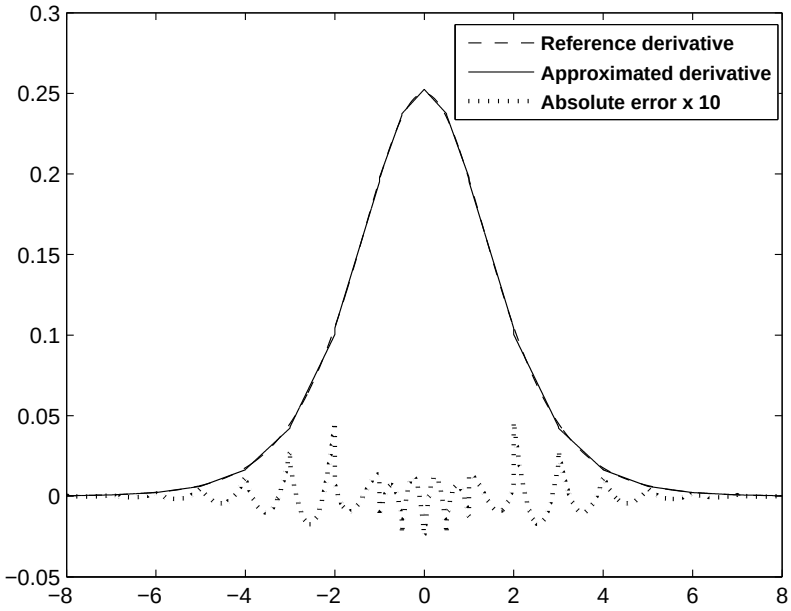

Figure 4.9: Approximated derivative sigmoid function

Table 4.3: Polynomial approximation of the sigmoid derivative function

Range	c'_1	c'_2	**Range**	c'_1	c'_2
[-8,-7]	0.00053	0.00456	$]0,\frac{1}{2}]$	-0.03029	0.25249
]-7,-6]	0.00143	0.01092	$]\frac{1}{2},1]$	-0.07727	0.27515
]-6,-5]	0.00387	0.02552	]1,2]	-0.09101	0.28625
]-5,-4]	0.01023	0.05735	]2,3]	-0.05718	0.21474
]-4,-3]	0.02578	0.11968	]3,4]	-0.02578	0.11968
]-3,-2]	0.05718	0.21474	]4,5]	-0.01023	0.05735
]-2,-1]	0.09101	0.28625	]5,6]	-0.00387	0.02552
$]-1,-\frac{1}{2}]$	0.07727	0.27515	]6,7]	-0.00143	0.01092
$]-\frac{1}{2},0]$	0.03029	0.25249	]7,8]	-0.00053	0.00456

4.3.4 Hyperbolic tangent approximation

Considering a rapid convergence to the asymptotes, the study of the hyperbolic tangent function is limited to the range [-4, 4] where it is continue and integrable. Moreover the function is a symmetric curve with respect to the point (0,0), the study can be reduced to [-4, 0]. This range is subdivided into equal subintervals of amplitude h=0.5 and each subinterval [a,b] is approximated by a straight line:

$$T_{approx}(x) = t_1 x + t_2 \tag{4.20}$$

where t_1 and t_2 are unknown quantities to be determined. The 0.5 amplitude of the subinterval is connected to the relationship between hyperbolic tangent and sigmoid, that is always limited in the range [-8,8] (Section 4.3.5). Applying the areas method in each of them we achieved:

$$t_1 = \frac{2}{(b-a)}[\frac{ln(\frac{e^{2b}+1}{e^{2c}+1})}{(b-c)} - \frac{ln(\frac{e^{2c}+1}{e^{2a}+1})}{(c-a)}] \tag{4.21}$$

$$t_2 = \frac{(b+c)}{(b-a)(c-a)} ln(\frac{e^{2c}+1}{e^{2a}+1}) - \frac{(c+a)}{(b-a)(b-c)} ln(\frac{e^{2b}+1}{e^{2c}+1}) - 1 \tag{4.22}$$

Table 4.4: Summary of absolute errors in the published literature for the hyperbolic tangent

Method	**Input range**	**Absolute Error**	
		$E_{ave}\%$	$E_{max}\%$
From Sigmoid	**[-4,4]**	**0.0011**	**0.0158**
Area Method	**[-4,4]**	**0.0011**	**0.0192**
LUT 9 bits [87]	[-4,4]	0.0040	0.0365
RALUT 9 bits [87]	[-4,4]	0.0089	0.0357
LUT 10 bits [87]	[-4,4]	0.0020	0.0180
RALUT 10 bits [87]	[-4,4]	0.0057	0.0178
Lin Scheme-1 [86]	[-4,4]	0.0078	0.0430
Lin Scheme-2 [86]	[-4,4]	0.0041	0.0220
Vassiliadis [76]	]-4,4[	0.0050	0.057

In the Table 4.4 the comparison with the state of the art and with the sigmoid method (Section 4.3.5) is reported. In the Table 4.5 the polynomial approximation is reported.
The approximated hyperbolic function is shown in Fig.4.10 and its absolute error in Fig. 4.11. In this case it is not realistic to consider the relative error because the (4.12) and the (4.13) are discontinue in

Table 4.5: Polynomial approximation of the hyperbolic tangent function and c subdivision points

Range	t_1	t_2	c	**Range**	t_1	t_2	c
[-4,-3.5]	0.0023015	-0.9902171	-3.720	]0,0.5]	0.9606801	5.90E-05	0.003
]-3.5,-3]	0.0062475	-0.9765672	-3.219	]0.5,1]	0.6138875	0.166917	0.686
]-3,-2.5]	0.0168879	-0.9450791	-2.719	]1,1.5]	0.290379	0.4803449	1.209
]-2.5,-2]	0.0452475	-0.875324	-2.218	]1.5,2]	0.1182683	0.7321556	1.716
]-2,-1.5]	0.1182683	-0.7321556	-1.716	]2,2.5]	0.0452475	0.875324	2.218
]-1.5,-1]	0.290379	-0.4803449	-1.209	]2.5,3]	0.0168879	0.9450791	2.719
]-1,-0.5]	0.6138875	-0.166917	-0.686	]3,3.5]	0.0062475	0.9765672	3.219
]-0.5,0]	0.9606801	-5.90E-05	-0.003	]3.5,4]	0.0023015	0.9902171	3.720

a neighborhood of O(0,0), therefore we just consider the minimum of the maximum of the absolute error.

Similarly the method has been applied to approximate its derivative. The detailed piecewise approximation results of the derivative function are shown in Table 4.6 and in Fig. .

4.3.5 Hyperbolic tangent approximation by the sigmoid

As reported in Section 4.3.4 the study of the hyperbolic tangent function is limited to the range [-4, 4] where it is continue and integrable. Based on the equation (4.2), the equation (4.3) can be rewritten as:

$$T(x) = 2S(2x) - 1 \tag{4.23}$$

The equation (4.23) shows the relationship between hyperbolic

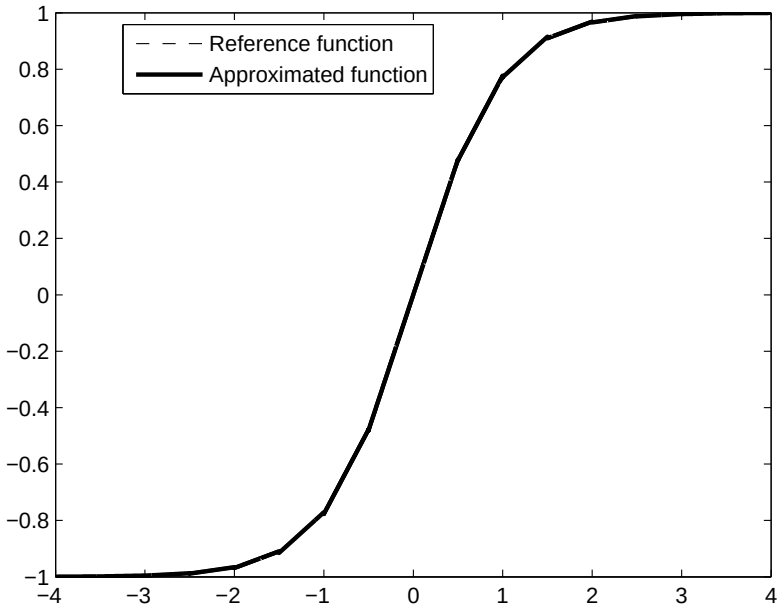

Figure 4.10: Approximated hyperbolic tangent with areas method

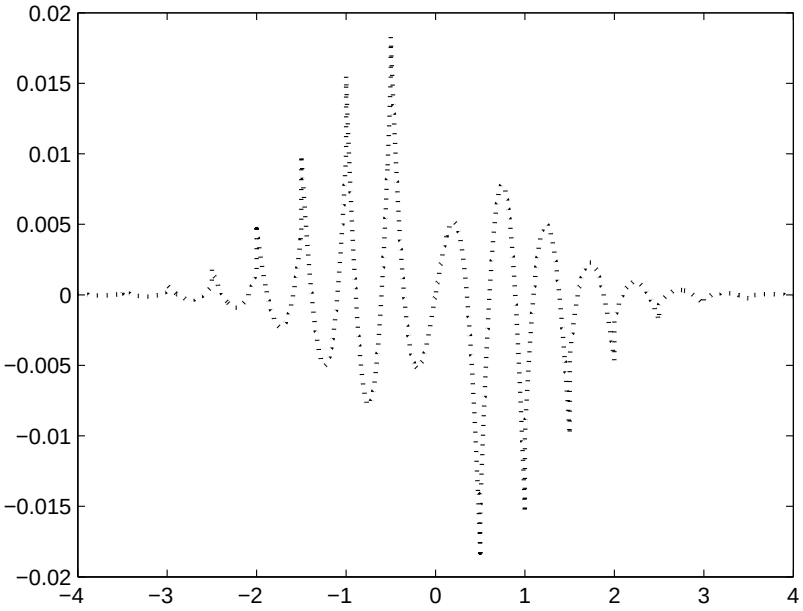

Figure 4.11: Absolute error of hyperbolic tangent with areas method

Table 4.6: Polynomial approximation of the hyperbolic tangent derivative function

Range	t'_1	t'_2	**Range**	t'_1	t'_2
[-4,-3.5]	0.00425	0.01825	]0,0.25]	-0.24239	1.00997
]-3.5,-3]	0.01152	0.04368	]0.25,0.5]	-0.61838	1.1007
]-3,-2.5]	0.03098	0.10209	]0.5,1]	-0.72815	1.14506
]-2.5,-2]	0.08189	0.22942	]1,1.5]	-0.45772	0.859256
]-2,-1.5]	0.20633	0.47883	]1.5,2]	-0.20633	0.47883
]-1.5,-1]	0.45772	0.859256	]2,2.5]	-0.08189	0.22942
]-1,-0.5]	0.72815	1.14506	]2.5,3]	-0.03098	0.10209
]-0.5,-0.25]	0.61838	1.1007	]3,3.5]	-0.01152	0.04368
]-0.25,0]	0.24239	1.00997	]3.5,4]	-0.00425	0.01825

tangent and sigmoid. Since in each unitary subinterval]a,b], from the (4.16), $(2x\epsilon]a,b])$,

$$Sapprox(2x) = c_1(2x) + c_2 \tag{4.24}$$

it follows that, in each subinterval]a/2,b/2], of amplitude h=0.5:

$$T_{approx}(x) = 2S_{approx}(2x)-1 = 2(c_1(2x)+c_2)-1 = 4c_1x+(2c_2-1) = t_1x+t_2 \tag{4.25}$$

where

$$t_1 = 4c_1;\ t_2 = 2c_2 - 1 \tag{4.26}$$

Therefore, to calculate the approximated value of the hyperbolic tangent in a generic point x, in the range]a/2,b/2], we substitute in

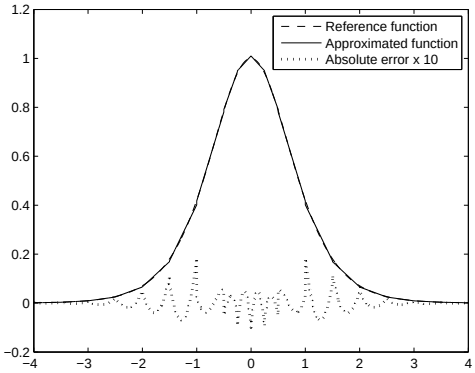

Figure 4.12: Approximated derivative hyperbolic tangent function with areas method

(4.26) the values of c_1 and c_2 calculated from the sigmoid approximation into]a,b] thus obtaining t_1 and t_2. Finally, substituting the values t_1 and t_2 into (4.25) we obtain Tapprox(x). Analogously for hyperbolic tangent derivative, we can observe that:

$$D(T(x)) = \frac{1}{\cosh^2(x)} = \frac{4}{(e^x + e^{-x})^2} = \frac{4e^{-2x}}{(1 + e^{-2x})^2} = 4S'(2x) \tag{4.27}$$

where S'(2x) is the sigmoid derivative calculated in $2x$. Analogously, in each subinterval]a,b], ($2x\epsilon]a, b]$), the sigmoid derivative is approximated by the function below:

$$S'_{approx}(2x) = c'_1(2x) + c'_2 \tag{4.28}$$

thus obtaining, in each subinterval]a/2,b/2]:

$$T'_{approx}(x) = 4S'_{approx}(2x) = 4(c'_1(2x) + c'_2) = 8c'_1x + 4c'_2 = t'_1x + t'_2 \tag{4.29}$$

where

$$t'_1 = 8c'_1;\ t'_2 = 4c'_2 \tag{4.30}$$

Finally, to calculate the approximated value of hyperbolic tangent derivative in a generic point x, inside the range]a/2,b/2], we can proceed as above. Moreover, observing that (4.25) and (4.29) can be implemented in hardware only with shift and subtraction by using the sigmoid function. The approximated hyperbolic function is shown in Fig.4.13 and its absolute error in Fig. 4.14.

The performance of the method is compared with the main approximation studies (Table 4.4). The approximated hyperbolic derivative function is shown in Fig.4.15

4.3.6 VLSI Circuit Design

As reported in Section 4.3.1, many works have been produced on hardware realization of sigmoid function. Most of them however, concerns the ASIC technology, in which the minimization of the area as a cost metric, in terms of gate-count, cannot be transferred in the context of FPGAs. In fact, FPGAs have a pre-existing structure of reprogrammable functional elements (CLB or LE), which cannot be described as equivalent gate-count. However, we focus the implementation on FPGA, with the advantage of high speed of prototyping and low cost of this platform. An interesting solution is in [64], which specifically adapted for implementation on

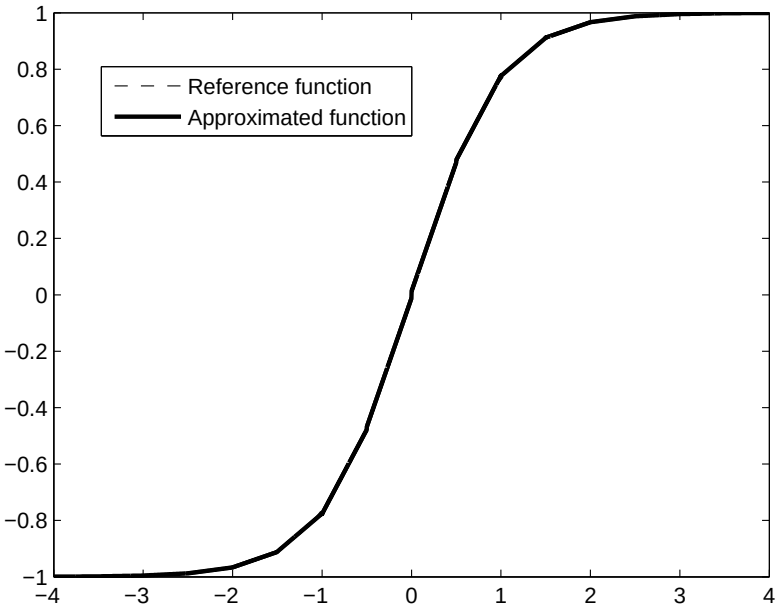

Figure 4.13: Approximated hyperbolic tangent

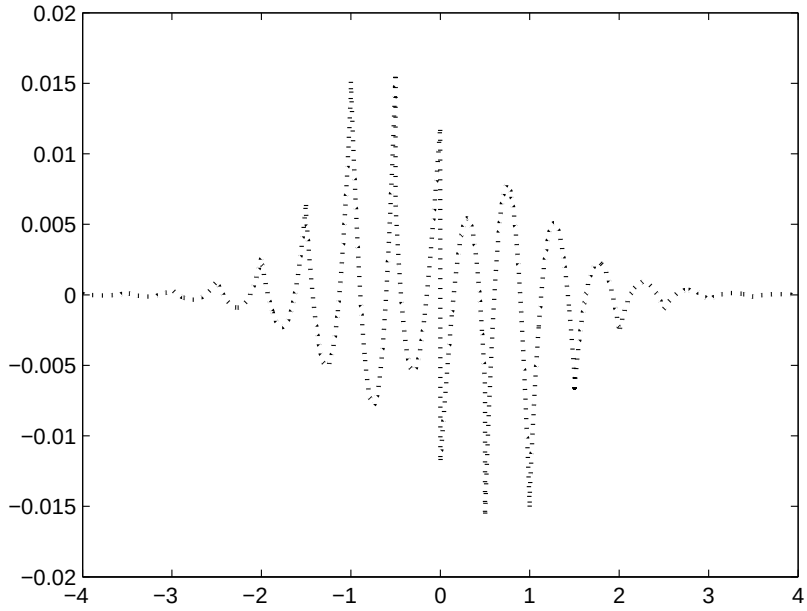

Figure 4.14: Absolute error of hyperbolic tangent

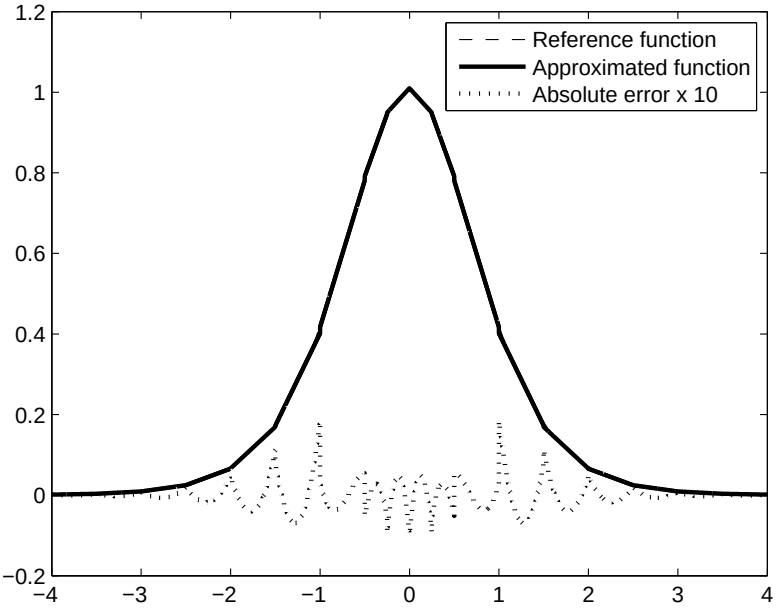

Figure 4.15: Approximated derivative hyperbolic tangent function for sigmoid derivative

FPGAs. However, the target is a small FPGA (Altera APEX II) and the claim made in [64] that "assuming generic multipliers is not realistic", is out of date, because the current FPGAs provide a large number of multipliers [77]. Starting from this hypothesis, a different basic approach, validated by the work [89], is to combine low-order polynomials and small LUT. Applying this technique, we take into account the main cost metrics, reported in [64] [67], to obtain an efficient activation function on FPGA. Assuming that an implementation in finite arithmetic machine (FXP) inevitably introduces approximation and rounding errors, we explored, as shown in Table 4.7, the different levels of error, that can be obtained for the sigmoid with our PWL approximation, in function of the number of precision bits. In this case, to compare our study with [77], the error values are obtained for 2^{10} uniformly sampled points. We addressed a further issue: the convergence of a neural network in relation to the precision bits of weights and activation function. In this context, some studies [70] [72] report the minimum number of precision bits to train via back propagation a multi-layer perceptrons (MLP). So, in order to achieve a complete study, we decided to test our activation function, as reported in [72], through the classic problem of the exclusive OR (XOR), which is described in detail in [90]. The topology selected for this problem is (2,2,1), two inputs, two hidden neurons, and a single neuron in the output layer. To achieve convergence data, weight sets are randomly generated and training phase is performed using a learning rate of 0.1 for 10000 epochs. Among these sets, 200, which give adequate convergence (<20% root-mean square error), are selected and the convergence is assessed by fixed point software training for different precision levels. The function

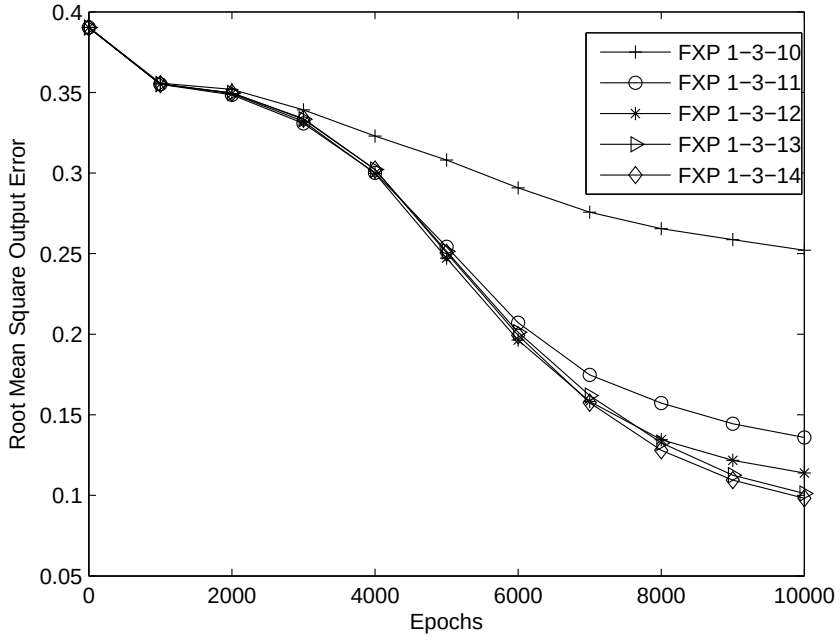

Figure 4.16: Average convergence of network using FXP formats

has a maximum usable input integer range of 8 and a maximum usable output range of 1. We plot the average error data for the 200 simulations in each FXP numeric formats in Fig. 4.16 which shows that to have convergence of the networks a minimum of 11-bit precision is needed.

Based on these results, it is proposed a 2's complement 4.12 fixed point minimal representation (1 bit sign, 3 integers bits and 12 fractional bits) for inputs and 0.12 fixed point representation (0 integers bits and 12 fractional bits) for outputs.

Table 4.7: Error of the sigmoid approximation in function of the precision bits compared with Bajger

Precision	$E_{aveAbso} \times 10^{-3}$	$E_{MaxAbso} \times 10^{-3}$	$E_{aveRel} \times 10^{-2}$	E_{MaxRel}
	Our / Baiger[77]	Our / Baiger[77]	Our / Bajger[77]	Our / Bajger[77]
Exact	1.068 / 1.199	7.470 / 8.616	1.349 / 1.547	$7.503\text{x}10^{-2}$ / $6.147\text{x}10^{-2}$
8 bits	3.820 / 4.272	1.080x10 / 1.081x10	9.178x10 / 9.271x10	1.185x10 / 1.186x10
10 bits	1.609 / 2.216	7.515 / 8.199	2.053x10 / 5.401x10	3.283 / 9.736
12 bits	1.140 / 1.348	7.449 / 8.688	6.252 / 7.027	1.267 / 1.184
14 bits	1.095 / 1.221	7.510 / 8.626	3.034 / 1.991	$4.790\text{x}10^{-1}$ / $2.720\text{x}10^{-1}$
16 bits	1.069 / 1.201	7.480 / 8.611	1.390 / 2.037	$1.354\ \text{x}10^{-1}$ / $1.631\ \text{x}10^{-1}$

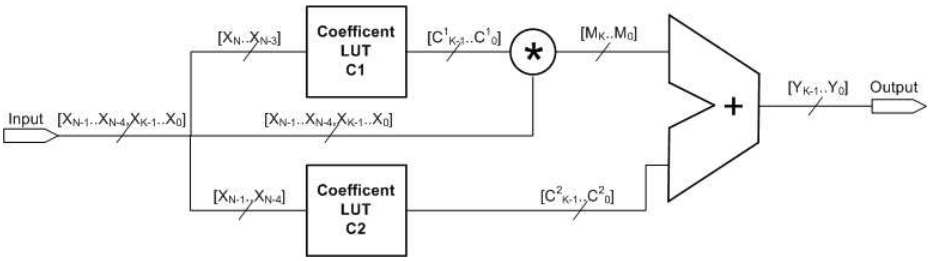

Figure 4.17: Activation Function-Scheme 1.

The FPGA used to implement the activation function is the Virtex-4 that has got Xtreme DSP slices for arithmetic [91]. The proposed digital circuit implementation (lookup-multiply) shown in figure 4.17 requires two lookup tables, ROMs or RAMs, which hold the constants c_1 and c_2 (t_1 and t_2 etc.) for each subinterval.

The most significant input bits, are used as index selector into the two LUTs. The scheme 4.17 can be used for sigmoid, hyperbolic tangent and their derivatives.
Although the claim to avoid multiplication is not necessary thanks to the existence of a large number of multipliers in current FPGA, in the scheme reported in (Fig. 4.18) a hybrid solution, which is valid for sigmoid activation function, that bypasses the multipliers is proposed. The LUT contains the coefficients c_1 multiplied by the integer part and one fractional bit of the input. The remain of fractional part is multiplied by coefficients c_1 approximated using powers of two (the multiplication is performed by only shifts). In this case, the error is limited because it is introduced by the less significant part of the multiplication.

In the scheme reported in (Fig. 4.19) a flexible digital imple-

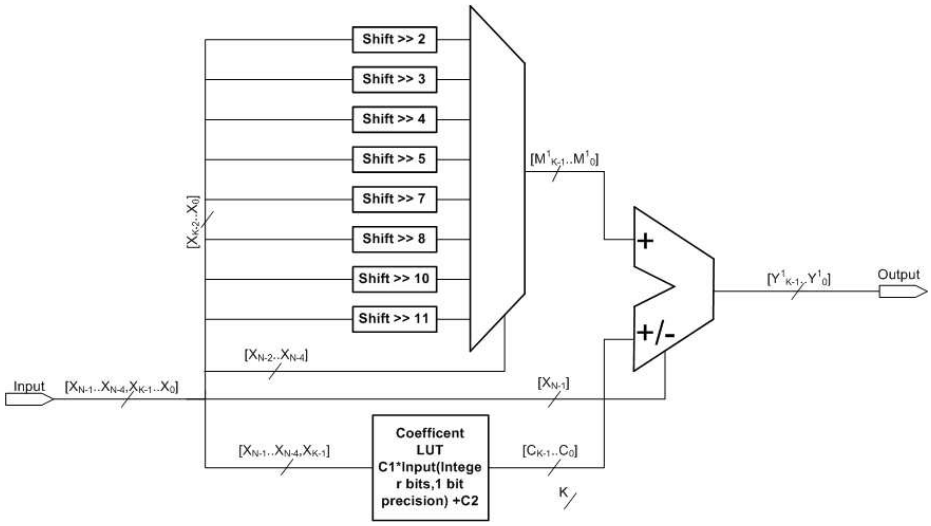

Figure 4.18: Activation Function-Scheme 2.

mentation of the activation function is showed where the hyperbolic tangent is generated by sigmoid coefficients. Therefore, the last implementation can be used as a flexible neuron in general purpose networks.

4.3.7 Sigmoid digital implementation comparative evaluation

To evaluate the different implementations reported in the state of the art, we have proposed a realization also on Altera APEX II, already used for the comparison study in [1]. Comparing the different works, also in relation to the cost metrics previously considered, the studies for which the convergence of the network was not guaranteed, are

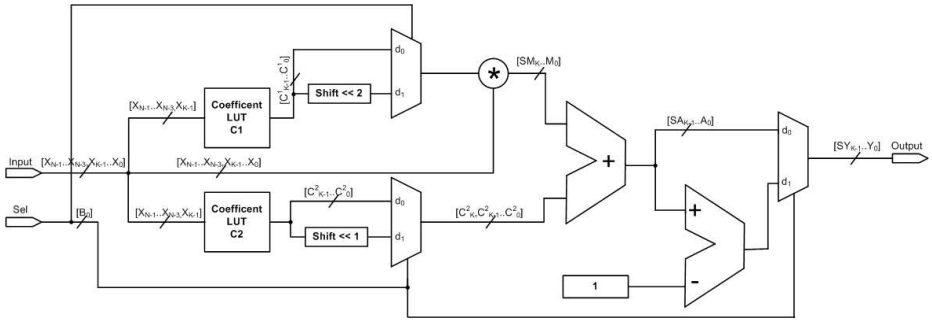

Figure 4.19: Scheme 3: Generator of Hyperbolic Tangent from Sigmoid function.

discarded. In figure 4.20 on the horizontal axis we report the frequency and on the vertical axis the error. The graph puts in evidence that in the pipeline implementation of the scheme reported in the Figure 4.17, the working frequency is significantly increased, highlighting the high quality of the solution compared to the others. The solution given by [64] is slightly better in terms of error. However this approach is not readily extensible to medium or high precision [77]. In our comparison work, we do not consider the work of Faiedh [84] because it is implemented in floating point and the recent study proposed in [77] which does not provide the approximation coefficients. In figure 4.21 we report, again by way of comparison, on the horizontal axis the number of logic elements and on the vertical axis the frequency. This graph shows, once again, that our solution is the best in terms of work frequency in spite of the larger area occupied. In addition, we take into account the high number of precision bits used, in relation to the study of convergence and the error study

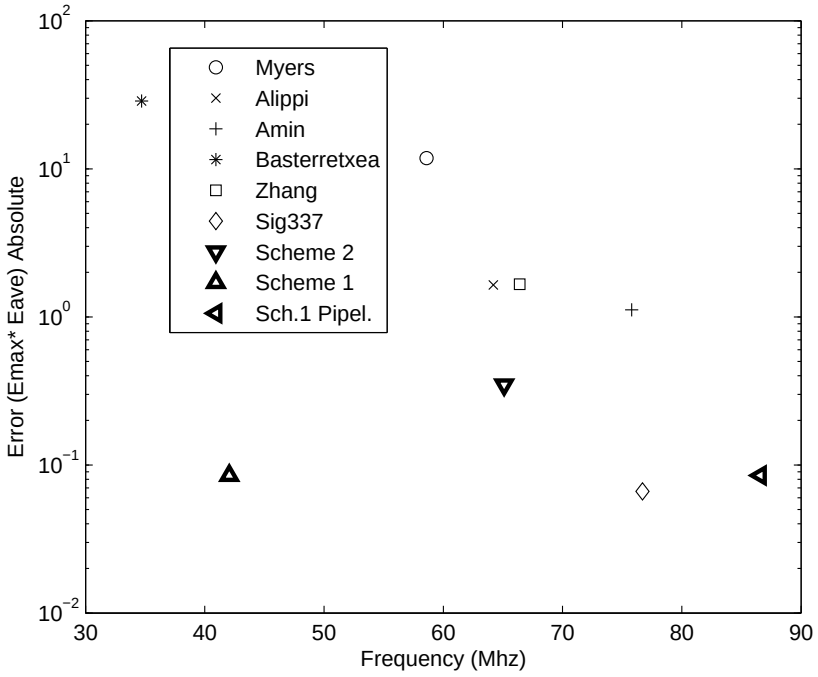

Figure 4.20: Comparative evaluation of hardware implementations on EP2A15F672C7 by error and frequency

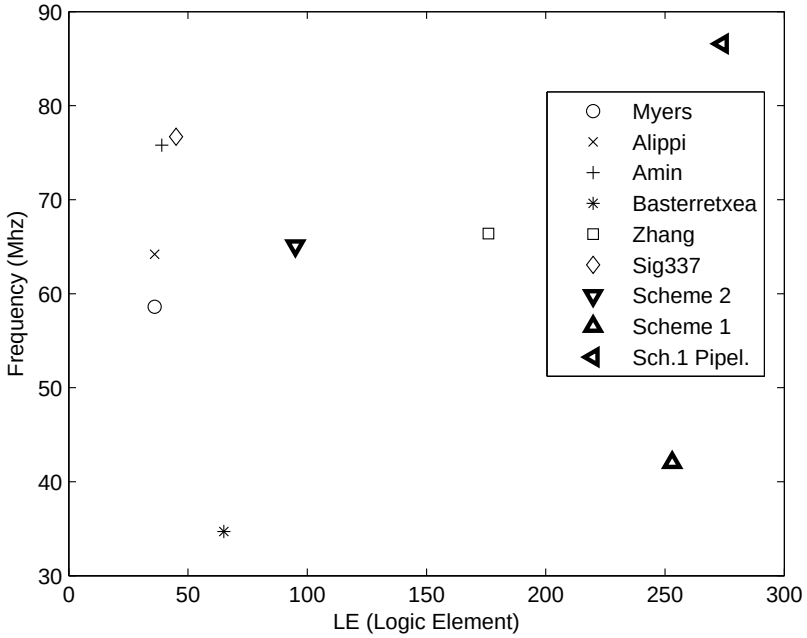

Figure 4.21: Comparative evaluation of hardware implementations on EP2A15F672C7 by logic element and frequency

made in the Section 4.3.6.

4.3.8 Topology of the ANNs in Hardware

Although FPGAs are widely accessible reconfigurable circuits (RC), there are two obstacles in their usage in evolutive systems. The first obstacle is that in current FPGAs the reconfiguration system is not sufficient for evolutive systems. In fact, internal reconfiguration inside the FPGA can configure the programmable elements of the same FPGA (e.g the Internal Configuration Access Port in Xilinx FPGA) (ICAP) it is still too slow for evolvable hardware applications. The second obstacle is that FPGAs can be reconfigured externally with the result of an interruption of the system operation during the download of the bitstream.
An efficient and fast reconfiguration subsystem is a desired feature for building evolvable hardware applications, so in order to overcome the problem of slow reconfiguration, Sekanina [92] has developed Virtual Reconfigurable Circuits (VRCs). Based on VRC, I propose a Virtual Reconfigurable Network (VRN) (Fig. 4.22) where the routing circuits are created using multiplexer.

The configuration of the router is codified in a gene together with the weights of each neuron. The configuration is storage in a shared memory (Fig.4.22) increasing the speed performance of the network. A customized software Genetic Algorithm (GA) is used to evolve the topology of the network. The network is tested on the logical-XOR problem with a population of 100 chromosomes. The selection operator is a modification of a roulette wheel selection. In fact, this selection method may cause noise by selecting individuals that are

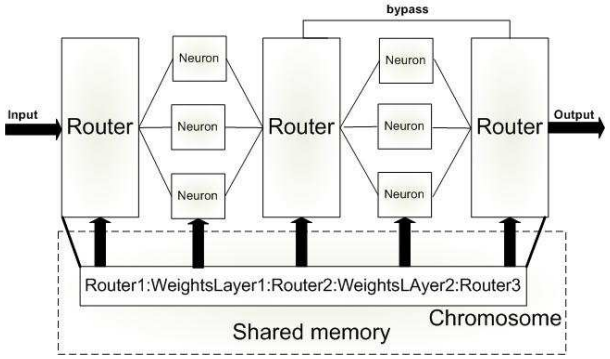

Figure 4.22: The Virtual Reconfigurable Network (VRN)

almost similar or by selecting bad ones. Also, there is more chance of losing good genes and becoming trapped in the local optima. These problems are prevented by making each individual's roulette wheel different from the ordinary method that uses one common roulette wheel for all individuals. The individual roulette wheels contain f_i permission space and $1 - f_i$ non-permission space, and these factors help determine the choices where f_i is the fitness function calculated as the difference between the output of the networks and the XOR output. The process is repeated until the number of selected individuals is reached. Figure 4.23 shows the convergence of all solutions. It is known that the solution 2-1-1 does not converge in the XOR problem according to the results. Figure 4.24 shows the number of individuals for each kind of solution that evolve over time.

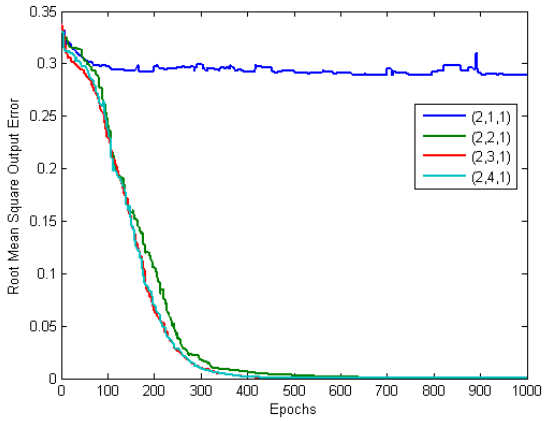

Figure 4.23: Convergence of different species corresponding to different neural networks topology

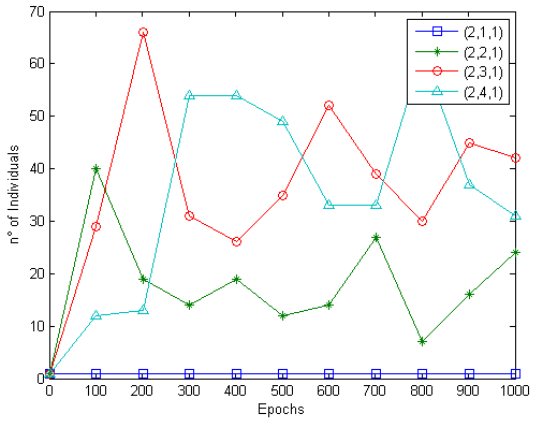

Figure 4.24: Number of individuals for different species

4.4 Results and Discussions

The area method adopted provided a very simple methodology to approximate the sigmoid function, hyperbolic tangent and their derivative. Moreover, the choice of the point c ϵ]a, b[coinciding with the minimum point of the curve of the maximum relative error or absolute error, has provided a systematic method. The obtained results are the lowest in terms of absolute and relative errors. In fact, as reported in Table 4.2, referring to the approximation of the sigmoid, the area method has provided the lowest absolute mean error and the lowest absolute maximum error, and regarding the relative error, the lowest average error. The relative maximum error is comparable with the results of Bajger [77].
The PWL approximation of the hyperbolic tangent, extrapolated by the mathematical relationship between this function and the sigmoid (Table 4.4), is the best in terms of average and maximum absolute error. Similar considerations can be made for the derivative of the hyperbolic tangent, extrapolated by the mathematical relationship between this function and the sigmoid derivative.
The PWL approximations of the hyperbolic tangent and its derivative, directly obtained by the area method, are comparable with the results obtained by the mathematical relationship with sigmoid.
The sigmoid digital implementation has been compared with the state of the art. Scheme 1 shows in Figure 4.20 the best performance with respect to the other implementations except for Tommiska [64], which is comparable in terms of error. However, the work [64] reports a very low number of precision bits in input. To obtain adequate convergence of a network, we have to consider the study

[70] as well as the recent study [72]. Therefore, based on our convergence study we have considered more bit precision of the state of the art to the detriment of the performance in terms of gate complexity (Fig.4.21). The flexible topology proposed has been tested on the logical-XOR problem, demonstrating the capability to reconfigure the architecture of the network. The GA guarantees the evolution of different topologies.

Chapter 5

Eye Tracking System

There is a growing body of research that makes use of eye-tracking technology to study attention disorders and visual processing in Autistic Spectrum Disorders (ASD). Atypical gaze patterns were already described for individuals with ASD when presented with social scenes and faces [54] [55]. For instance, in social environments, individuals with ASD show reduced eye-region fixation time in favour of an increased focus on mouths and objects [56]. Atypical gaze fixation patterns in ASD have been also reported during the observation of still photos of faces. More specifically, it has been demonstrated that there exist clear differences in how individuals with ASD scan and process both familiar and unfamiliar facial images [54]. Interestingly, children with ASD are reported to increase visual attention to the demonstrator's face when his actions were non-meaningful [55]. These results seem to reinforce, in a more fine-grained manner, the same disorders in face processing that were behaviorally observed in ASD. These disorders may include failures

in coordinating visual attention with others and the abnormal identification of emotional expressions and mental states. On the other hand, prominent research groups have demonstrated that individuals with high-functioning autism differ from controls in how they process faces, but only when those faces are incorporated into social-dynamic stimuli, hypothesizing also a possible relation with insufficient mirror neuron activity. Crucial issues to be investigated involve the replications of these findings making use of carefully designed experimental paradigms, together with a more fine-grained investigation of the processes underlying specific autism deficits (such as imitation, emotional responsiveness and social processing). To this aim, the gaze tracking technology should provide a suitable solution, which can be unobtrusively and ecologically used in social experimental paradigms. In this chapter, the HATCAM is presented, which is an innovative wearable device that was specifically designed to investigate early attention disorders in infants. The HATCAM device was designed to meet the following requirements:

1) wearability;

2) minimal obtrusiveness;

3) eye tracking and pupillometry capabilities;

4) lightweight (less than 100 g);

5) wireless communication.

6) attractive and aesthetic;

7) real-time elaboration;

Moreover, an innovative emotion classification study based on gaze pattern is proposed in Section 5.8. This classification could be used to support the emotion classification centered on physiological signals.

5.1 State of the arts

Eye Gaze Trackers (EGTs) have been and are currently used in many research areas, such as marketing and advertising, as well as in human factor engineering to evaluate computer interfaces and web sites. Moreover, EGTs have been studied as input devices for computer interfaces especially for people with disabilities. A specific field of application, on which researchers currently focusing their efforts, is the study of oculomotor characteristics and abnormalities, and their relation to cognition and mental states. EGTs rely on an analytical relationship between relative eye movements and the point of gaze over time, and all of the current systems that can be found in literature are differentiated by the method used to calculate eye movements. There are several eye-movement measurement methods based on mechanical or optical referenced objects mounted either on the scleral contact-lens (Fig.5.1) or directly on the eye (e.g. search coil), which results to be very precise although invasive.

Other methods, based on ElectroOculoGram (EOG), [93], [94], are affected by several disadvantages such as obtrusiveness, low resolution, drift, noise, EMG artifacts, in addition the discomfort for the patient who is limited in his normal activities, [95]. A specific class of EGTs that try to overcome these drawbacks is based on Video-

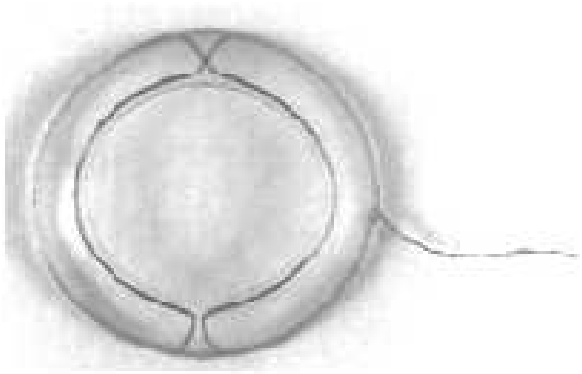

Figure 5.1: Contact lens system

OculoGraphy (VOG), which is a method for tracking eye movements by digitally processing video images of the eyes. Pupil position and iris landmarks are detected by means of image processing algorithms and are used to calculate eye rotation angles and as well as its center. Currently, EGTs can be split into two different classes, remote and wearable. Remote eye-tracking systems require the user to keep his or her head still, thus making the systems unsatisfactory for prolonged use in interactive applications, even though they achieve an optimal accuracy, especially for clinical usage [96]. The second class, commonly referred to as head-mounted EGTs, uses eye tracking systems, i.e. cameras or other equipments, mounted directly on the user's head. The development of these wearable EGTs (Fig. 5.2 and Fig. 5.3) have opened up new scenarios where the user is free to move thereby overcoming the spatial limitations imposed by the remote systems, [97]. In addition, their use outside controlled environments allows for investigating eye-movement measurements

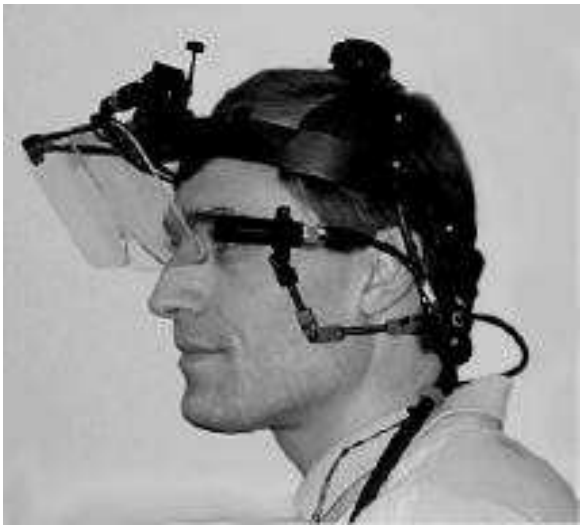

Figure 5.2: Wearable EGT system based on IR mirror [99]

during natural tasks, e.g. driving cars, [98], hand washing, [99], walking through a room cluttered with obstacles, etc.

On the other hand, the extended workspace of EGTs introduces some difficulties in estimating the pupil center during illumination changes, as well as requires less obtrusiveness, lightness and is not accessible to a large audience, e.g. children. At present, eye localization, independent of illumination, is still a crucial issue. Current systems, indeed, are constrained to be used in controlled environments where illumination is kept constant. However, in the last few years, many efforts have been made to overcome this limitation. One of the most significant approaches is based on eye detection and tracking by using active InfraRed (IR) illuminators. In particular, spectral

Figure 5.3: The wearable eye tracker: headgear and backpack, headgear with an omnidirectional vision sensor, and detail of backpack [100]

(reflective) properties of the pupil, under near-IR illumination, are exploited to maximize the image-contrast between the pupil and the background (Fig. 5.4). If the IR illumination is aligned properly with the pupil, the bright pupil is obtained (Fig. 5.4d); otherwise you get a dark pupil (Fig. 5.4c).

Based on this principle, several tracking techniques, [102] and [103], have been developed as well as some commercial eye trackers have been produced, [104] and [105].
For example, OpenEyes, developed by Dongheng et al. [106] is a low cost, head-eye tracker system composed of two CCD cameras mounted on a pair of safety glasses. In this system, one camera captures an image of the eye while the other captures an image of the scene. An InfraRed (IR) LED is placed in axis with the eye-camera producing an illumination which allows, together with an ad-hoc algorithm based on Random Sample Consensus (RANSAC)

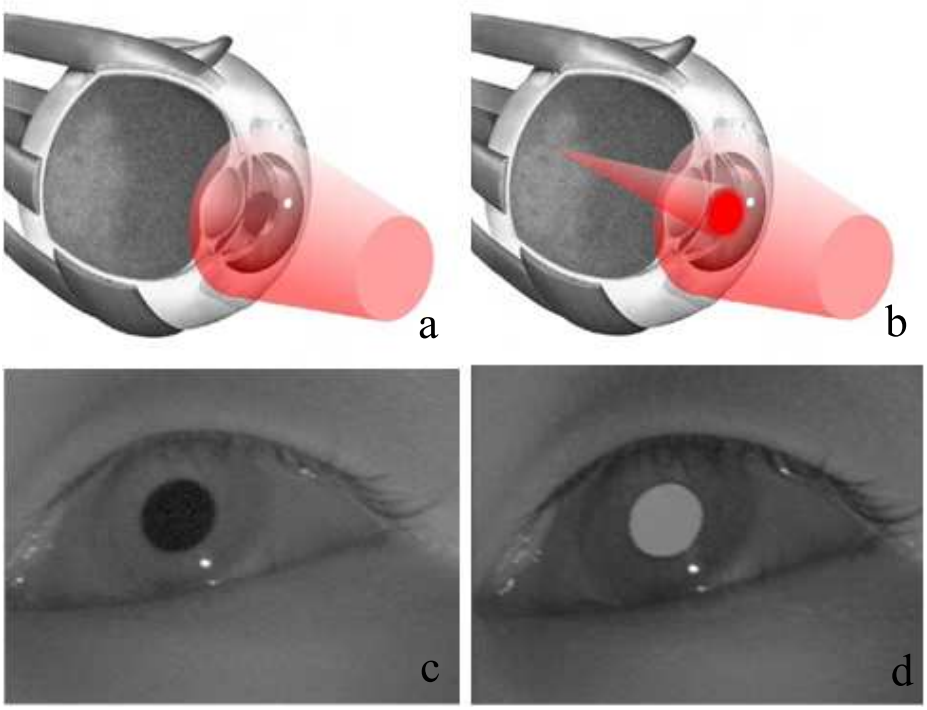

Figure 5.4: a) An infrared source illuminates the eye. b) When aligned properly, the illumination beam enters the eye, retro-reflects off the retina and back-illuminates the pupil. c) Dark pupil. d) Bright pupil [101].

paradigm, for the discrimination of the pupil from the rest of the eye [107]. Unfortunately the RANSAC is an iterative algorithm. Consequently, the system is not able to perform real-time applications [108]. Moreover, the use of two cameras introduce a latency between the scene and eye camera complicating the calibration; in this case the synchronization of both videos is achieved introducing a flash light, which is uncomfortable. Another eye tracker based on glasses and an IR illuminator has been introduced by Babcock [101]. This EGTs is based on an IR mirror, an IR camera, and an IR illuminator, but the mirror appears obtrusive especially when the subjects are children, and the use of the IR camera and illuminator forces the system to be used in a controlled environment. Moreover, such a system is heavy and cannot be used to examine people who wear eyeglasses or contact lenses. Babcock also introduced an eye tracker based on glasses [109] which overcomes the problem of holding the head in a stationary mode during the calibration phase by introducing laser projection systems, which provided a calibration plane fixed to the head.

Even though uniform and controlled infrared illumination eliminates uncontrolled specular reflection and is not perceivable by the user, it does not permit the natural aversion response which protects the eyes against retinal injuries when viewing very bright light sources, [110]. Even though low-power infrared LEDs are generally employed to avoid injuries, the irradiance level must be kept to less than $10\text{mW}/cm^2$ for not chronic IR exposure in the range of 720-1400 nm, [111]. Moreover, the use of IR illuminator systems, especially for high-sensitive subjects, e.g. children, can produce reddening and lachrymation. Furthermore, in subjects with eyeglasses the lens dis-

turbs the infrared light thus showing very weak pupils. Finally, a large variation of bright light sources can produce a diminished image of the pupil or even its disappearance. These limitations in eye tracking methods using IR illuminators impose stable lighting conditions, and therefore a restriction of the fields of application.
An interesting prototype has been proposed by Ryan [112], which is a glasses system able to track eye movements under variable lighting conditions through the tracking of the limbus , but that was not suitable for children. Franchak et. al. [113] introduced very lightweight systems that are suitable for children based on two light cameras and an IR LED to produce a reference point on the pupil, but the system was not wireless. The LASA of Ecole Politechnique de Lausanne proposed Wearcam [57], a new eye tracking system specifically designed for children. The system is based on a wearable wireless camera located on the forehead and a mirror in order to obtain in the same frame the eyes of patient and the scene.

Therefore, it is very important to develop an EGT able to robustly and accurately track eyes under several illumination conditions and during their changes, to calculate the gaze-point in real-time and lightness in order to increase its acceptability for children. In the next sections a detailed description of HATCAM is provided describing the software and hardware solutions for obtaining an innovative device according to the requirement reported in the introduction. Particular attention is paid to describing the solution in order to overcome the limitations found in the other devices such as robustness under illuminations conditions, real-time elaboration, lightness and the latency between the scene and eye camera.

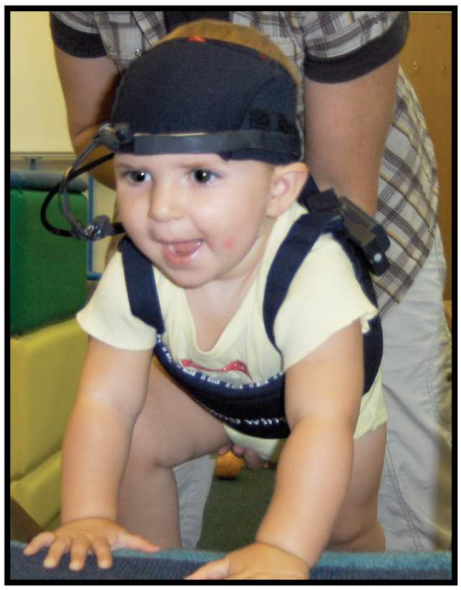

Figure 5.5: Infant wearing the eye-tracker [113].

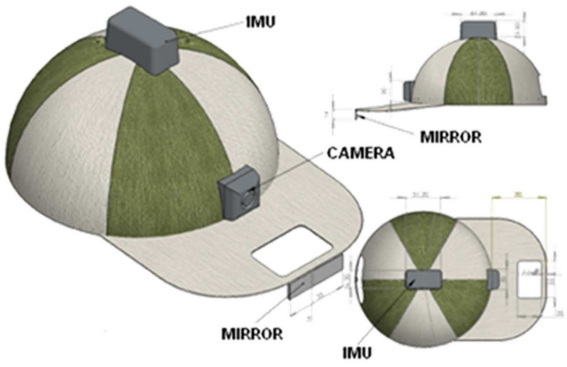

Figure 5.6: Version I: "Baseball-like" hat configuration.

5.2 HATCAM

HATCAM is a new wearable and wireless eye tracking system, which can be tailored to both adults and children. Two configurations are presented: "baseball-like" hat (see fig.5.6) and head band (see fig. 5.7). In Figure 5.8 a child is wearing HATCAM.

They are comprised of only one lightweight camera able to capture simultaneously the eyes of the subject and the scene in front of him by means of a mirror. The system can be used in indoor and outdoor environments. In detail, the system is comprised of a wireless CMOS camera (CP294) being light in weight ($20g$), small in size ($2x2x2cm$), and having an A/V transmitter of up to $30m$ of distance. The camera has a resolution of $628x586$ pixels with $F2.0$, $D45°$ optic, and 25 frames per second (f.p.s.). The original lens of the camera was removed and substituted with a wide-angle-lens without an IR filter. This operation allows for enlarging the view angle and

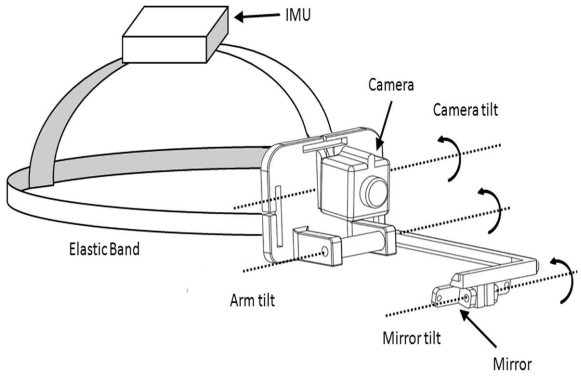

Figure 5.7: Version II: "Head band" configuration.

Figure 5.8: A control subject wear HATCAM.

acquiring infrared components, which emphasize the contrast between pupil and iris. This system is able to capture simultaneously the visual scene in front of the subject and the position of his eyes without latency,. This is achieved using a mirror ($4x0.6cm$) placed on a shaft linked to the head (see fig. 5.7). The tilt and shaft of the mirror and camera orientation can be tailored to the forehead profile of the user (see fig.5.7).
Moreover, the system is also equipped with a wireless IMU for both monitoring the head movements during natural activities and taking into account the "Vestibulo-Ocular Reflex" (VOR) generated by the interaction between the visual and motor-kinesthetic system (for maintaining a stable gaze). This reflex consists of an eye rotation for head movement compensation to stabilize vision [114, 115]. Unfortunately, the IMU is sensitive to external magnetic fields, e.g. mobile phones. It provides three rotation euler-angles of the head (see Fig.5.9) at a frequency of 20 Hz, and is mounted on a strap over the top of the head close to the azimuth rotation center (see fig. 5.9).

5.3 Eye tracking Method

This section deals with the processing techniques used to detect the center of the eye and how its movements are mapped into the image plane. VOG method involves visible spectrum imaging. This technique is a passive approach that captures ambient light reflected from the eye. The mounted camera is modified to acquire also the IR components of natural light. Therefore, the system keeps the advantages

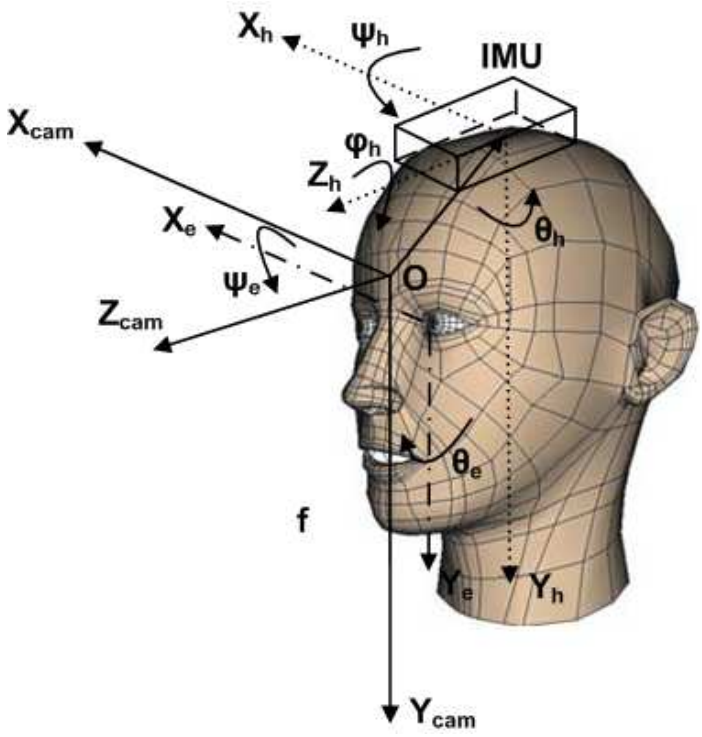

Figure 5.9: Representation of the head along with IMU. In the figure X_e, Y_e are the axes of the eye; X_{cam}, Y_{cam}, Z_{cam} are the axes of the camera; X_h, Y_h, Z_h are the axes of the center of the head where the IMU is placed; ψ_h, θ_h, ϕ_h are the euler-angles of the head rotation while ψ_e, θ_e, are the euler-angles of the eye rotation.

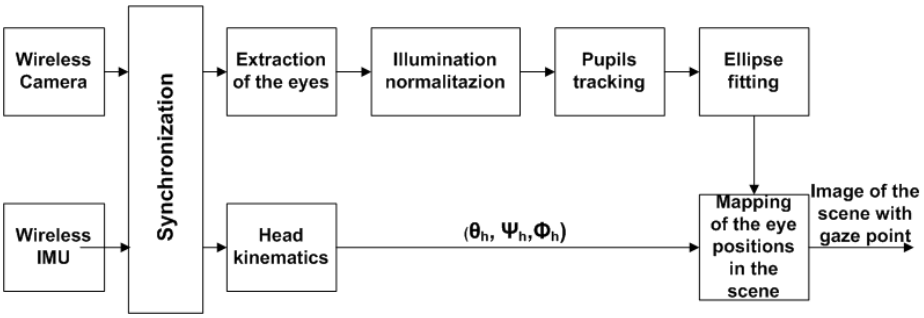

Figure 5.10: Block diagram showing all of the algorithmic stages of the processing of eyes and outside scene.

of IR illumination in increasing the contrast between pupil and iris, and at same time preventing any possible injuries due to artificial IR illuminators. Indeed, as the bandwidth includes both visible light and IR components, no illuminators are needed. Figure 5.10 shows the block diagram of the image processing. In this diagram, the image of the eyes together with the scene are inputs of the processing chain. More specifically, the first processing block implements the eye extraction algorithm; the second applies the photometric normalization algorithm of illumination; and the third block extracts the pupil contour and implements the fitting algorithm. At this point the center of the eye is detected. The last block consists of an appropriate function, which maps the eye center and movements into the image plane. This function is generally named "$mappingfunction$". In the next sections, each block is described more in detail.

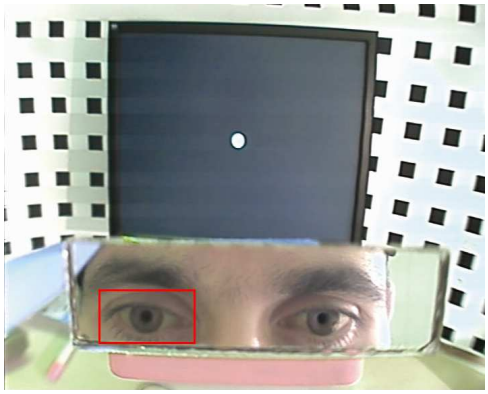

Figure 5.11: Example of a single frame captured by the camera. The rectangular area marked by red represents the ROI.

5.3.1 Eye Extraction

Figure 5.11 shows how the HATCAM is able to acquire simultaneously the eyes of the user and the scene in front of him using the mirror. The eye extraction procedure consists of visual inspection of the first video frame, in which a rectangular area including the eye is freehand selected (see fig. 5.11 and 5.12). This region is called the Region Of Interest (ROI). Since eye tracking is a head-mounted system, the ROI does not change throughout the experiment. In addition, only the red-image-component is converted into gray scale and is used as an input to the other processing blocks (see fig. 5.12), as this component is specifically helpful in enhancing the contrast between the pupil and background.

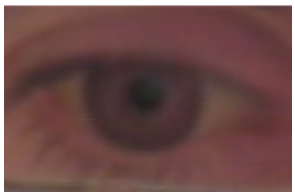

Figure 5.12: Red component of the ROI.

5.4 Photometric normalization techniques

As reported in the state of the art (Section 5.1), the main difficulties in the wearable EGT is the robust estimate of the pupil center during illumination changes. An illumination normalization technique has recently been proposed in order to solve the brightness variation issue in the different fields of application such as face recognition [116] or automotive scenarios [117] introduced to the EGT field.
The purpose of the illumination normalization is to reduce or eliminate some variations in the captured eyes due to different conditions of illumination. It normalizes and enhances the eye image to improve the recognition performance of the system. The photometric normalization consists of removing the mean of the geometrically normalized image and scaling the pixel values by their standard deviation, estimated over the whole cropped image [118]. All of the photometric normalization techniques applied in this work are based on either Homomorphic Filtering or Histogram Equalization. Six out of the seven explored photonormalization techniques are based on the Retinex theory (from the words "*retina*" and "*cortex*", suggesting that both the eye and brain are involved in the processing) devel-

oped by Land [119]. This theory is based on the color constancy assumption, which ensures that the perceived color of objects remains relatively constant under varying illumination conditions. Land and his colleagues assume that the stimulus is not the result of the light source and surface reflectivity only, but that the visual system processes the stimulus, integrating the spectral radiance and generating a ratio of integrated radiance of any region of the scene with that of the brightest region. This stimulus is called *lightness*. This model eliminates the effect of a non-uniform illumination and is completely independent of any apriori knowledge of the surfaces reflectance and light source composition.

Two major assumptions underlie this theory:

- The human visual system performs the same computation independently of the color channel;
- In each channel, signal intensity is proportional to the product of the illumination source and the surface reflectance, which is determined by the object characteristics.

According to this theory, the image intensity $I(x, y)$ can be simplified and formulated as follows:

$$I(x, y) = R(x, y)L(x, y) \tag{5.1}$$

where $R(x, y)$ is the reflectance and $L(x, y)$ is the illuminance at each point (x, y). The luminance L is assumed to contain low frequency components of the image while the reflectance R mainly includes the high frequency components of the image. This assumption can be easily understood if considering that R generally varies much faster

than L in most parts of the image with a few exceptions, e.g. shadow boundaries, where L changes remarkably. In addition, in a real world scenario, the illumination can dynamically change much more than the reflectance. In conclusion, there is a widely accepted statement about human vision, which assumes that human eyes respond to local changes in contrast to global brightness levels.
A comparative performance evaluation of the most relevant and widely used photometric normalization techniques is proposed in Section 5.6 in order to asses the best technique for HATCAM system. More specifically, eye gaze accuracy, pupillometry, and algorithm execution time have been evaluated in three different illumination conditions: darkness, laboratory and sunlight conditions. Below a detailed description of the seven photonormalization techniques, used in this comparative study, is reported.

Single Scale Retinex (SSR)

The latest version of the Land Retinex theory was implemented by Jobson as a Single Scale Retinex (SSR), [120]. The main idea of the algorithm is to process the image through a class of center surround functions where each output value of the function is determined by the corresponding input value (center) and its neighborhood (surround). The center is defined as each pixel value and the surround is a Gaussian function. The mathematical form of the SSR is given by:

$$SSR(x,y) = log(I(x,y)) - log[F(x,y) * I(x,y)] \qquad (5.2)$$

where $SSR(x, y)$ is the Retinex output, $I(x, y)$ is the input image, and $[F(x, y) * I(x, y)]$ is a convolution product between $I(x, y)$ and $F(x, y)$. This latter function is a simple linear filter with Gaussian kernel:

$$F(x, y) = ke^{\frac{-r^2}{\sigma^2}} \tag{5.3}$$

where $r = \sqrt{x^2 + y^2}$, σ is the standard deviation of the filter, empirically determined, and controls the amount of spatial detail that is retained in terms of pixels, and k is a normalization factor that keeps the area under the Gaussian curve equal to 1.

Multi Scale Retinex (MSR)

Although the SSR algorithm produces good results with a properly selected Gaussian filter, it is still limited by an important weakness: at large illumination discontinuities caused by strong shadows that are casted over the scene, halo effects are often visible in the computed reflectance, [121]. To avoid this inconvenience, Jobson extended the SSR algorithm to a multi-scale form, where Gaussian filters with different widths are used, and the output combines different implementations of distributed SSR algorithms to compute the final illumination invariant image representation. In detail, the MSR output is simply a weighted sum of the outputs of several different SSR outputs. Mathematically,

$$R_{MSRi} = \Sigma_{n=1}^{N} w_n R_{ni} \tag{5.4}$$

where N is the number of scales, R_{ni} is the ith component of the nth scale, R_{MSRi} is the ith spectral component of the MSR

output, w_n and is the weight associated with the nth scale. The only difference between $R(x, y)$ and $R_n(x, y)$ is that the surround function is given by

$$F_n(x, y)) = ke^{\frac{-r^2}{\sigma_n^2}} \tag{5.5}$$

Single scale selfQuotient Image (SQI)

The SQI algorithm is an extension of the Quotient Image (QI), [122], [123]. QI is a concept introduced by Riklin-Raviv and Shashua, [124]. It is based on the Lambertian model where the image can be described by the product of the albedo and the cosine angle between a point light source and the surface normal:

$$I(x, y)) = \rho(x, y)\mathbf{n}(x, y)^T \cdot \mathbf{s} \tag{5.6}$$

where $\rho(x, y)$ is the albedo,$\mathbf{n}(x, y)^T$ is the surface normal (shape) of the object (same for all objects of the class), and $\mathbf{s}$ is the light source direction, which may vary arbitrarily. QI of two objects belonging to the same class is defined as the ratio between the albedo of each object. The albedo represents the diffuse reflectivity or reflecting power of a surface. It is defined as the ratio of reflected radiation from the surface to incident radiation upon it. Being a dimensionless fraction, it may also be expressed as a percentage and is measured on a scale from zero for no reflecting power of a perfectly black surface, to 1 for perfect reflection of a white surface. The SQI algorithm is defined as follows:

$$Q(x, y)) = \frac{I(x, y)}{F * I(x, y)} \tag{5.7}$$

where $Q(x, y)$ is the result of the algorithm, $I(x, y)$ is the input image as defined in QI, and F is a smoothing filter. Unlike the QI, the SQI has several advantages, among which the most relevant for our purposes is that no image training processes are required which implies that it can be used as a pure pre-processing algorithm. A crucial issue to be carefully addressed in this technique is the filter size. If the scale is too small, the filtered image will be close to the original image, but if the filter is too large, the filtered image will be more or less constant, and the input image will be normalized with its mean value.

Multi Scale selfQuotient image (MSQ)

The Multi Scale selfQuotient image technique exhibits similarities to MSR, but unlike MSR, this technique uses an anisotropic filter for the smoothing operation. Moreover, MSQ allows for overcoming the limitation of the filter size by using several scales, as in the MSR, in order to achieve more robust results.

Discrete Cosine Transform(DCT)

Chen et al. [116] proposed an alternative technique to normalize illumination. This approach is employed to compensate for illumination variations by truncating the low frequency components of the discrete cosine transform (DCT) in the logarithm domain.

WAvelet-based Normalization algorithm (WAN)

Du and Ward [125] presented a preprocessing technique based on the wavelet transform. The authors proposed to apply the histogram equalization to the sub-band image generated by the so-called approximation wavelet-coefficients and to emphasize the remaining sub-bands generated by the detail-wavelet-coefficients. Through the histogram equalization step, the image contrast is improved while the second step enhances the edge information. The final compensated image is obtained by simply employing the inverse wavelet transform on the modified coefficient sub-bands.

Normalization Histogram (NH)

Other photometric normalization techniques can be also based on Histogram Equalization. Histogram equalization is usually achieved by equalizing the histogram of the image pixel gray-levels in the spatial domain so as to redistribute them uniformly. Struc, [126], proposed the possibility to replaced it with an arbitrary distribution such as the normal, the lognormal and the exponential.

5.4.1 Pupil tracking

The pupil tracking algorithm extracts the contour of the pupil exploiting the higher contrast of the pupil than the background due to the IR components of the natural light. Figure 5.13 shows the algorithm block diagram. More in detail, the first block applies a binarization of the image, obtained by using a threshold, which can be chosen according to two different criteria. Fig. 5.14 reports the his-

togram of the eye, i.e. the distribution of the image pixel vs. gray levels from 0 to 255. The threshold should divide the histogram into two groups of pixels having only two levels of gray; the zero level should group all of the pixels belonging to the pupil, whereas level 255 should identify the background. The first criterion implies choosing the threshold as the absolute minimum value in the range comprised between the two highest peaks of the eye histogram as reported in Fig. 5.14. The latter criterion implies taking the threshold as the level of gray at which the cumulative sum of pixels, starting from 0, achieves 3 % of the total pixels of the image. This value was experimentally estimated in order to discriminate the pupil at best. An example of the binarization process is reported in Fig.5.15. The group of second, third and fourth blockssh constitutes the preliminary pupil contour detection. More specifically, the second block identifies two sheafs of lines starting from the middle points of the vertical sides of the image, with an angular aperture of 30°. As a result of the binarization process, the image borders are expected to belong to the background, therefore the starting point of each line has a value of 255 in terms of gray level. Analogously, the pupil is expected to be placed roughly in the middle of the image (this is assured by the an accurate freehand selection of the ROI). When each line encounters a dark pixel along its path, the latter can be thought to belong to the contour of the pupil. The third block calculates the centroid of these points. The fourth block removes all of the outliers, since these points are very far from the centroid with respect to the large point density (pupil edge). At this stage the set of the obtained points constitutes a large-grain approximation of the contour. The fifth block identifies a sheaf of lines starting from the

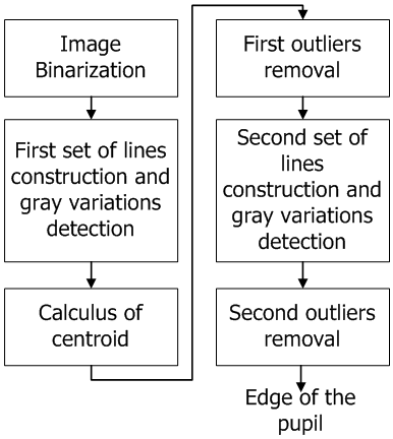

Figure 5.13: Block diagram of the pupil tracking algorithm

centroid with an angular aperture of 360°, and detects all discontinuities, but now from black to white. The sixth block is another removal operation of the outliers. The result of this algorithm is a set of points constituting the pupil edge. This set will be the input of the fitting algorithm (see 5.16).

5.4.2 Ellipse fitting

The ellipse fitting algorithm is implemented for pupil contour reconstruction and for detecting the center of the eye. The ellipse is considered as the best geometrical figure representing the eye, since the eye image captured by the camera is a projection of the eye in the mirror. According to the ellipse construction, it can be expressed by an implicit second order polynomial, being a central conic (with $b^2 - 4ac < 0$), such as:

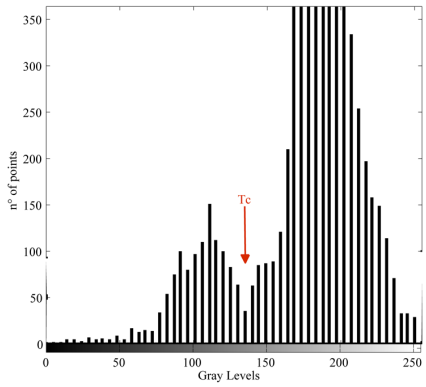

Figure 5.14: Example of the histogram of the eye: T_c refers to the threshold identifying the eye and the sclera region

Figure 5.15: Example the eye image after the binarization process.

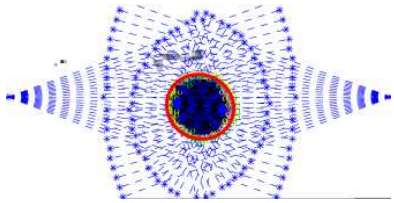

Figure 5.16: Application of the pupil tracking algorithm. More specifically, in black there are the set of points of the eye including the outliers; in blue the various line constructions are represented; in orange the pupil contours are highlighted.

$$F(x, y) = ax^2 + bxy + cy^2 + dx + ey + f = 0 \tag{5.8}$$

Ellipse fitting algorithms present in the literature can be divided into two broad techniques: the clustering/vot-ing (CV) and the least square (LS) techniques. The first one uses two main approaches such as RANSAC and Hough Transform. The RANSAC technique is extremely robust, but it is time-demanding and memory-consuming, [108], and as an iterative algorithm it does not have fixed computation time therefore not suitable for real-time applications. The Hough Transform suffers from sensitivity limitations due to the presence of spurious and blurred peaks, [127], and even though many efforts were made for computational cost reduction, the algorithm still seems to be excessively resource consuming for real-time machine vision, [128]. The LS method is based on finding a set of parameters that minimize the distance between the data points and the ellipse. According to literature, this technique fulfills the real time issue. One implementation of the LS technique has been intro-

duced by Fitzgibbon et al., which is a direct computational method (i.e. B2AC) based on the algebraic distance with a quadratic constraint, [129]. In this work, we use a custom B2AC algorithm, where a Gaussian noise is added for algorithm stabilization, [130], to calculate the center of the pupil (that coincides with the ellipse center), the axes dimensions, as well as the eccentricity.

5.4.3 Mapping of the eye positions in the scene

The mapping procedure associates the eye center position to the image plane of the scene, providing as a result the gaze point. An experimenter guides this procedure. Firstly, the camera is positioned to capture both the scene and the mirror. In detail, the tilt of the camera is adjusted as well as the tilt of the mirror shaft and the tilt of the mirror to reflect the eyes. Each participant is asked to look at some specific points of the screen (calibration process). These points are identified by coordinates $s_i = (x_{si}, y_{si})$ referred to as the image plane (i.e. the image plane captured by the camera), (see fig.5.11). The participants were instructed to keep their head as still as possible and to carefully look at each target point without blinking until looking at the next one. The mapping function receives as an input the center of the eye coming from Ellipse fitting block, and the coordinates of a point on the image plane. Mapping functions are quadratic polynomials defined as:

$$x_{si} = a_{11} + a_{12}x_{ei} + a_{13}y_{ei} + a_{14}x_{ei}y_{ei} + a_{15}{x_{ei}}^2 + a_{16}{y_{ei}}^2 \qquad (5.9)$$

$$y_{si} = a_{21} + a_{22}x_{ei} + a_{23}y_{ei} + a_{24}x_{ei}y_{ei} + a_{25}{x_{ei}}^2 + a_{26}{y_{ei}}^2 \qquad (5.10)$$

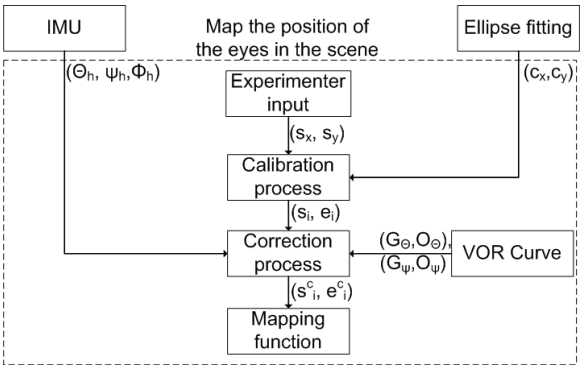

Figure 5.17: Block Diagram of the mapping function calculation process.

where x_{si}, y_{si} are the coordinates of the image plane (i.e. the coordinates of the point on the screen mapped into the image plane captured by the camera), and x_{ei}, y_{ei} are the coordinates of the center of the eye coming from the ellipse fitting block, referred to the image plane as well. The coefficients $a_{1,1-to-6}$, and $a_{2,1-to-6}$ are unknowns. Since each calibration point defines two equations, the system is over constrained with 12 unknowns and 18 equations, and can be solved using the Least Square Method (LSM). Figure 5.17 shows the process of calculating the mapping function. The process considers a correction algorithm responsible for compensating the head movement as described in Section 5.4.4.

5.4.4 Correction process

The main drawback of the calibration process consists in the subject not being able to hold his head firmly. Consequently, these

movements reduce system accuracy. Therefore, an ad-hoc correction process, based on data from IMU, is used for compensating the head movements. After the calibration, 9 target points $s_i = (x_{si}, y_{si})$ and 9 points related to eye positions $e_i = (x_{ei}, y_{ei})$ are acquired. Moreover, we also acquire Euler angles of the head (θ_h, ϕ_h and ψ_h see fig.5.18) for each target point, and they are used to correct the $s_i = (x_{si}, y_{si})$ points in the exact positions.

We define, according to Figure 5.18, $OX_{cam}Y_{cam}Z_{cam}$ the Cartesian system fixed to the camera, $O'x_iy_i$ the cartesian system on the image plane, f the focal distance, $c(x_c, y_c)$ the projection of the central point of the calibration plane on the image plane, $p(x, y)$ the projection of $P(x, y, z)$, which is a generic calibration point, on the image plane. Figure 5.20 shows a transverse view of the subject, the image plane and the calibration plane (or real plane), during the initial conditions.

In this case, the subject is aligned with the central point of the calibration plane (point number five, see fig.5.19)

During a positive head rotation of θ_h (see fig.5.21) (which is a reasonable head movement during the calibration process) the projection of P in the image plane results in p' instead of p. This means that an error is made during the calibration process and, consequently, is propagated in the gaze estimation. Therefore, by taking into account the acquired head rotations, it is possible oto correct the projection of P in the exact position by means of an algorithm based on some geometrical considerations reported in Figure 5.22. In the case showed in Fig.5.22 x'_p clearly results in being less than x_c and is not the only possible but others can be identified:

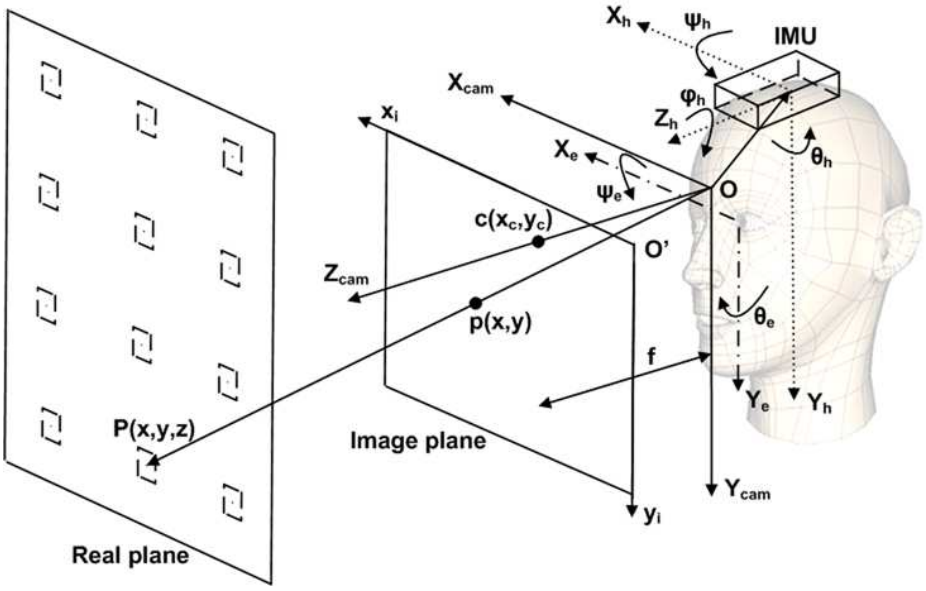

Figure 5.18: Representation of the head,IMU, image plane and real plane. In Figure X_e, Y_e are the axes of the eye; X_{cam}, Y_{cam}, Z_{cam} are the axes of the camera; x_i,y_i are the axes foe the image plane; X_h, Y_h, Z_h are the axes of the center of the head where the IMU is placed; ψ_h, θ_h, ϕ_h are the euler-angles of the head rotation while ψ_e, θ_e, are the euler-angles of the eye rotation; $c(x_c, y_c)$ is the projection of the central point of the calibration plane on the image plane; $p(x, y)$ the projection of $P(x, y, z)$, which is a generic calibration point, on the image plane.

- $x_{p'} > x_c$ and $(\theta' + \theta_h) < 90°$;
- $x_{p'} > x_c$ and $(\theta' + \theta_h) > 90°$;

and for $\theta_h < 0$, we have:

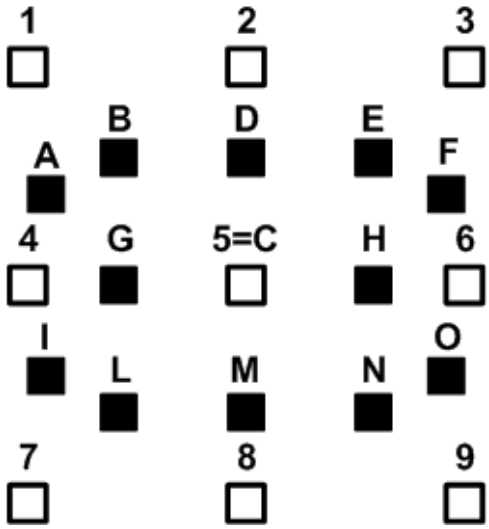

Figure 5.19: Calibration plane.

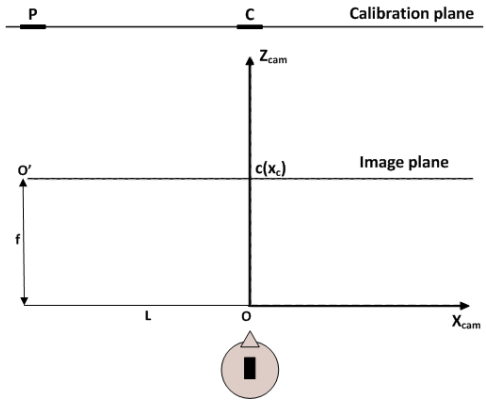

Figure 5.20: Initial condition of the calibration.

- $x_{p'} > x_c$ and $(\theta' - \theta_h) < 90°$;
- $x_{p'} < x_c$ and $(\theta' - \theta_h) < 90°$;

5.4 Photometric normalization techniques

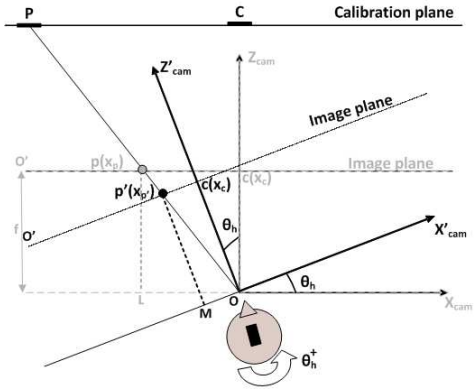

Figure 5.21: Representation of the positive θ_h rotation.

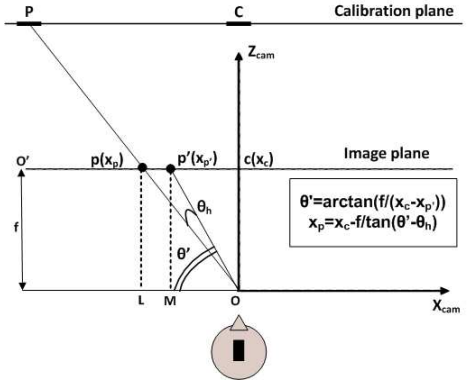

Figure 5.22: Representation of the final condition after the positive θ_h rotation

- $x_{p'} < x_c$ and $(\theta' - \theta_h) > 90°$;

Considering all of the cases mentioned above, the correction along the x axis can be summarized by the following equations:

$$\theta' = arctan(\frac{f}{|x_c - x_{p'}|}) \tag{5.11}$$

if $x'_p < x_c$ then

$$x_p = x_c - \frac{f}{tan(\theta' - \theta_h)} \tag{5.12}$$

otherwise

$$x_p = x_c + \frac{f}{tan(\theta' + \theta_h)} \tag{5.13}$$

The y_p corrections are the same as the x_p corrections using the angle ψ_h instead of θ_h, but the relations are inverted with respect to the sign of the angle. Therefore:

$$\psi' = arctan(\frac{f}{|y_c - y_{p'}|}) \tag{5.14}$$

if $y_{p'} < y_c$ then

$$y_p = y_c - \frac{f}{tan(\psi' - \psi_h)} \tag{5.15}$$

otherwise

$$y_p = y_c + \frac{f}{tan(\psi' + \psi_h)} \tag{5.16}$$

The correction by ϕ_h gives a contribution for both x and y coordinates. All cases for both $\phi_h < 0$ and $\phi_h > 0$ can be summarized by the following equations:

if ($x'_p < x_c$ and $y'_p < y'_c$) or ($x'_p > x_c$ and $y'_p > y'_c$) then

$$x_p = x_c - \frac{(x_c - x'_p)}{sin(\phi')} sin(\phi' + \phi_h) \tag{5.17}$$

$$y_p = y_c - \frac{(y_c - y'_p)}{sin(\phi')} cos(\phi' + \phi_h) \tag{5.18}$$

otherwise

$$x_p = x_c - \frac{(x_c - x'_p)}{sin(\phi')} sin(\phi' - \phi_h) \tag{5.19}$$

$$y_p = y_c + \frac{(y_c - y'_p)}{sin(\phi')} cos(\phi' - \phi_h) \tag{5.20}$$

Moreover, vestibulo-ocular reflex acts during all the cases described above where a head rotation is present. Therefore an eye repositioning is required and actuated by the VOR curves (Section). More specifically, the VOR contributions together with θ_h and ψ_h are used to correct the corresponding eye rotations according to the follow equations:

$$\theta'_e = \tilde{\theta}_e - \theta_h + \triangle(\theta_h) \tag{5.21}$$

$$\psi'_e = \tilde{\psi}_e + \psi_h + \triangle(\psi_h) \tag{5.22}$$

where θ'_e and ψ'_e are the corrected eye angles and $\tilde{\theta}_e$ and $\tilde{\psi}_e$ are the eye angles of a specific subject wearing the system. $\triangle(\theta_h)$ and $\triangle(\psi_h)$ are the decomposition of ϕ_h, which is the head rotation around the z axis.

5.4.5 Vestibulo-ocular calibration curve

Vestibulo-ocular calibration curves are needed to compensate the eye rotation with respect to the head movements. It is achieved by introducing a mathematical model for transforming the eye rotations, expressed in degrees, into movements of the eye center along with the axes (vertical and horizontal), expressed in pixels, while the head is rotating. This compensation is useful for removing the effect of the physiological Vestibulo-Ocular Reflex (VOR), which is responsible for the compensatory eye rotations in response to head motion, as sensed by the vestibular organs in the inner ear [131]. More specifically, during head rotations, VOR rotates the eyes in opposite directions for maintaining the gaze stable [114]. A linear fitting is applied to both rotations for extracting gain and offset, as expressed by the formula:

$$\theta_e = G_{\theta_e} P_x + O_{\theta_e} \tag{5.23}$$

$$\psi_e = G_{\psi_e} P_y + O_{\psi_e} \tag{5.24}$$

5.5 Determination of VOR curve

Thirteen subjects were recruited of both genders with different skin tones, with 8 subjects having dark-coloured eyes and 5 having light-coloured eyes .The average age was 27.8. The experiment aims at showing that the VOR curves can be calculated only one time on the whole set of 13 subjects as an average of VOR for the rotations ψ along with x axis and for the rotations θ along with the y axis.

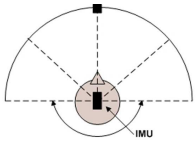

Figure 5.23: Experimental VOR calibration: Rotation around vertical axis (θ).

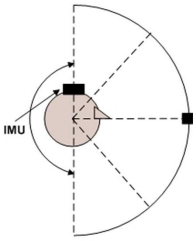

Figure 5.24: Experimental VOR calibration: Rotation around horizontal axis (ψ).

During the experiment, the subjects were asked to rotate the head about the horizontal and vertical axes keeping a fixed look on the point placed in front of them (see Fig.5.23 and Fig.5.24).

Finally, I obtained one gain and one offset for each subject, and the average gain of the whole set of volunteers was calculated, which remained the same for all subjects, while all offsets were discarded as they depend on how the system is worn. Figures 5.25 and 5.26 show the gains of the VOR curves (for x and y axes) calculated for each subject (continuous black line), and both average gain and standard deviation (dotted lines) of VOR curves for the first set of subjects. These average values are used as a gain correction to estimate each offset value of the specific subject in the second experiment.

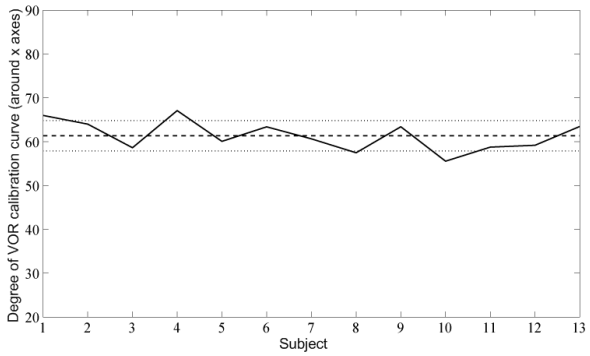

Figure 5.25: Average gain of VOR curve: Rotation around vertical axis (ψ).

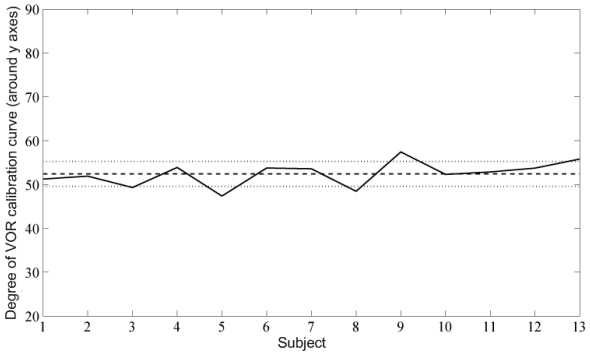

Figure 5.26: Average gain of VOR curve: Rotation around horizontal axis (θ).

Table 5.1 reports average gain and standard deviation of VOR for both vertical (θ) and horizontal (ψ) axes.

Table 5.1: Average Gain VOR

VOR around X	VOR around y
52.44±2.83	61.36±3.46

5.6 Experimental study of photonormalization technique

The aim was to increase the robustness of HATCAM at illumination variations comparing the performance of seven photonormalization algorithms on pupil area detection and gaze tracking under different illumination conditions. Ten subjects, nine males and one female, volunteered to participate in the experiment. Six subjects had dark-coloured eyes and 4 had light-coloured eyes. The average age was of 26.8 with a standard deviation of 1.5. The experiment was performed in three illumination conditions: darkness, laboratory and sunlight condition. In detail, the laboratory condition was achieved by white neon lighting equally distributed in the room with a power of 50 lumens. A dark condition is achieved by reducing the power of the illumination by 55%, and the sunlight condition is obtained by performing the experiment at 12:30 a.m. at the top floor of the university building in a room having two completely windowed walls. The experimental protocol was structured into three phases. In the first phase, the participant was instructed on the experiment modalities. In the second phase, the calibration procedure was performed, and finally the experimental test was carried out. The whole experiment lasted about 8 minutes; the first 4 minutes were used to

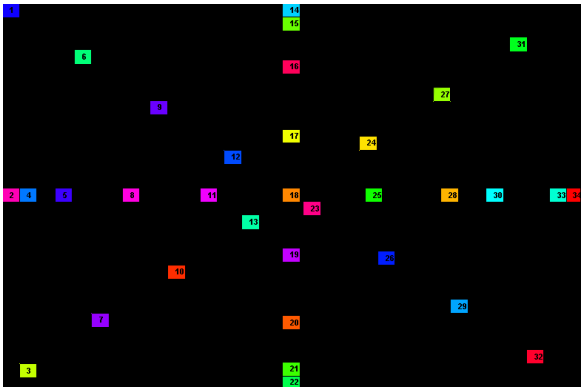

Figure 5.27: Example of the points showed on the screen for accuracy EGT system evaluation

describe the experiment and the remaining 4 to perform the calibration and testing part. More specifically, during the first phase all of the subjects were asked to sit on a comfortable chair in front of a screen at a fixed distance of 70 cm, while wearing the HATCAM system. The system was also equipped with a chin-support in order to avoid head movement. In the calibration procedure, the system was tailored (position and orientation of mirror and camera) to the user characteristic profile. During the experimental test, all subjects were asked to look at the point which appeared on the screen for 2.5 seconds each. Each subject was presented with 36 randomly-distributed points for each illumination condition (see Fig. 5.27) for a grand total of 108 points. The sequence of illumination conditions was randomly changed for each subject in order to avoid systematic errors.

The results are reported and discussed in terms of errors be-

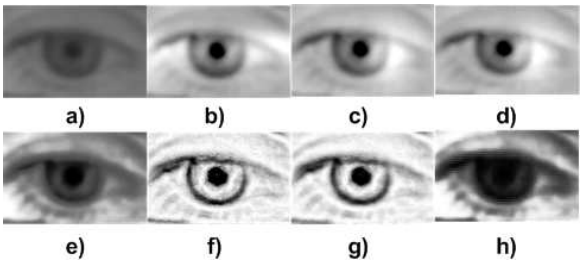

Figure 5.28: Example of all photonormalization techniques applied to the same gray eye image. In detail: (a) Original image; (b) Reconstructed image by applying DCT; (c) Reconstructed image by applying SSR; (d) Reconstructed image by applying MSR ; (e) Reconstructed image by applying NH; (f) Reconstructed image by applying SQI; (g) Reconstructed image by applying MSQ; (h) Reconstructed image by applying WAN;

tween estimated and real measurements. The performance evaluation of the different algorithms took into account the execution time as well. More specifically, Figure 5.28 shows an example of all of the photonormalization techniques applied to the same gray-level image of an eye. Although the algorithms provide results whose differences can be easily detected at a glance, a more quantitative comparison has to be made. A careful statistical analysis of errors calculated as the difference between real and estimated measurements for eye tracking and pupil area detection is performed. The errors, in terms of pixels, computed for all of the subjects and frames, are not normally distributed. In Figure 5.29, by way of example, the error histograms are reportedfor one algorithm and illumination condition.

This is also confirmed by the Lilliefors test, [132], which returns

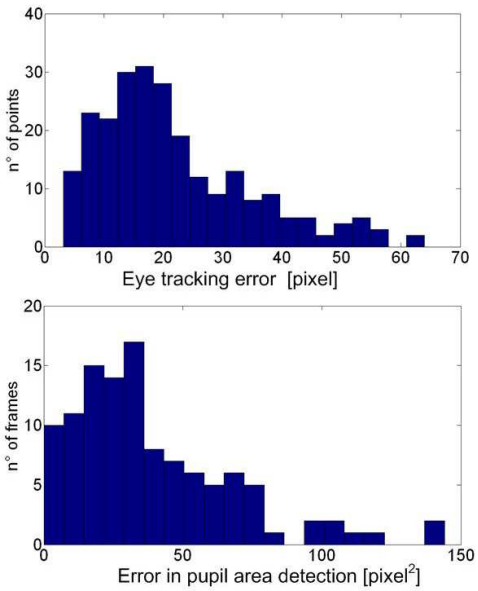

Figure 5.29: Example of the distributions of errors in eye tracking and pupil area detection.

a $p - value$ rejecting the null hypothesis of normality in all cases. Accordingly, we used the Kruskal-Wallis test, [133], which is a non-parametric, one-way analysis of variance by ranks for testing equality of population medians. Like most non-parametric tests, Kruskal-Wallis is performed on ranked data, so the measurement observations are converted to their ranks in the overall data set. This test does assume an identically-shaped and scaled distribution for each group, except for any difference in medians. The null hypothesis is stated as the probability that the samples come from identical populations, regardless of their distributions. In place of the mean distributions, the median as a measure of location is considered, usually taken when the distribution is skewed, such as in this case (see fig.5.29), [134]. Tables 5.2 and 5.3 show median and dispersion of each photonormalization technique in the three illumination conditions for both eye tracking and pupil area detection. The lowest values of the median are reported in bold. As it is well known from the statistical analysis, in order to evaluate performance of the seven photonormalization techniques, it is not sufficient to compare the means rather than medians (not being normal distributions) of the groups, but we have to assess whether the groups statistically come from different populations. Applying the Kruskal-Wallis test to all possible combinations among the seven techniques relative to eye tracking and for each illumination condition, four statistically equivalent classes of techniques are obtained, i.e. the null hypothesis in each class cannot be rejected. In Table 5.5, four different classes are identified for each illumination condition. The classes are sorted out by increasing median, therefore the best photonormalization techniques belong to class 1, as it exhibits the lowest median of error distributions. The

techniques in common with all the three illumination conditions are the NH and DCT. Similarly, the same methodology is applied for the pupil area error distributions, with results reported in Table 5.6. In this case, three classes for each illumination condition are identified. Following the same approach as above, the technique most similar to the three illumination conditions is the DCT. In addition to the error statistical analysis, the execution time of the seven algorithms under the three different illumination conditions is also considered. Even in this case the Lilliefors test was applied to the distributions of the execution time calculated for each frame of the video recordings, for all the photonormalization techniques and illumination conditions, giving as result that all the distributions are not normal. In order to compare the execution time, median and dispersion of execution time distributions for all the seven techniques were calculated and reported in Table 5.4. These values do not change under the three illumination conditions. The fastest algorithm is the NH. The choice of the most suitable technique depends on the need for an accurate rather than a fast algorithm. From Table 5.5 we can state that, for eye tracking in all the illumination conditions, DCT and NH are the best photonormalization techniques. Reading the first column of Table 5.6 we derive that for the pupil area detection in laboratory conditions DCT, SSR and MSR are the best techniques; in sunlight conditions DCT, SSQ, MSQ and WAN show the best results and in darkness conditions DCT, SSR, MSR and NH exhibit the lowest error. These results show that DCT is one of best techniques in pupil area detection, and it is also very effective for eye tracking. NH shows good results for eye tracking in all conditions and also in darkness conditions for pupillometry. The remaining techniques are,

in general, less effective under the three illumination conditions for both eye tracking and pupillometry. If I also take into account the execution time, the best technique was NH, which was executed in less than $1msec$, followed by SSR which was executed in less than $18msec$ (but effective only for pupillometry during laboratory and darkness conditions), and third is DCT, which was executed in less than $23msec$. The remaining techniques are slower.

5.7 Experimental validation of correction process

In this section using the DCT photonormalization technique and VOR curves previously determined the innovative algorithm introduced to correct head movement during the calibration procedure is validated. The experiment is conducted inside a cylinder of 1m ray (Fig.5.30) and is divided into two phases.

During the first phase the subjects were invited to align their eyes with the central point of the calibration plane (C point in the fig.5.19) called the initial condition, and were then asked to follow with their eyes the other calibration points indicated by the numbers, in an arbitrary order (Fig.5.19). Simultaneously, the experimenter clicked at the corresponding point seen on the image plane. In the second phase, the subjects were invited to follow the square indicated by a letter with their eyes . On average, about 5 minutes of data were recorded. During this experiment phase, the illumination conditions were modified to assess system accuracy in dark and sunlight conditions. Referring to the VOR contribution, the average

Table 5.2: Median and dispersion of eye tracking errors in pixels

	SSR	MSR	SQI	MSQ	DCT	WAN	NH
Laboratory	19.686±10.771	22.091±12.243	28.234±13.339	28.108±13.979	**16.214±7.016**	23.853±8.221	**17.114±7.364**
Sunlight	18.712±7.045	18.067±7.384	32.375±17.215	34.064±17.731	**16.522±6.641**	26.036±9.135	**16.053±6.235**
Darkness	23.037±12.908	21.415±11.870	29.052±13.874	34.368±19.000	**16.051±6.595**	22.504±5.643	**17.599±7.229**

Table 5.3: Median and dispersion of pupillometry errors in $pixel^2$

	SSR	MSR	SQI	MSQ	DCT	WAN	NH
Laboratory	**35.84±24.25**	**36.02±21.28**	36.23±26.49	35.68±23.71	**30.81±20.72**	47.54±26.26	39.09±15.66
Sunlight	48.88±23.12	49.37±21.06	**25.38±17.99**	**30.64±17.58**	**35.71±18.17**	**33.48±21.40**	76.34±20.69
Darkness	**45.97±24.77**	**46.98±26.07**	69.49±33.79	71.42±36.99	**41.63±26.20**	110.69±34.02	**37.55±16.67**

Table 5.4: Median and dispersion of the execution time [s]

SSR	MSR	SQI	MSQ	DCT	WAN	NH
0.0175±0.0011	0.0407±0.0018	0.3028±0.0199	1.3700±0.0496	0.0224±0.0019	0.0255±0.0050	**0.0098±0.0008**

Table 5.5: Eye tracking error classes

		Class 1	Class 2	Class 3	Class 4
Laboratory	Techniques	DCT,NH	SSR	MSR	SQI,MSQ,WAN
	p-value	0.6242			0.3399
Sunlight	Techniques	DCT,NH	SSR,MSR	WAN	SQI,MSQ
	p-value	0.5921	0.7577		0.9077
Darkness	Techniques	DCT,NH	SSR,MSR,WAN	SQI	MSQ
	p-value	0.3426	0.6688		

Table 5.6: Pupillometry error classes

		Class1	Class 2	Class 3
Laboratory	Techniques	DCT,SSR,MSR	NH,SQI,MSQ	WAN
	p-value	0.3548	0.9478	
Sunlight	Techniques	DCT,SQI,MSQ,WAN	SSR,MSR	NH
	p-value	0.9512	0.6143	
Darkness	Techniques	DCT,SSR,MSR,NH	SQI,MSQ	WAN
	p-value	0.1364	0.9400	

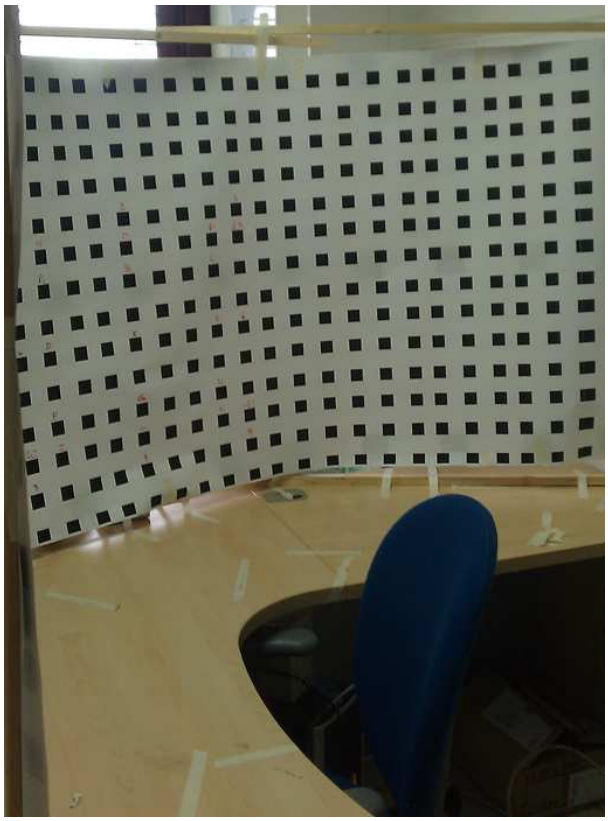

Figure 5.30: Subject emplacement: Cylinder around which the experiments are performed.

gain obtained in Section 5.5 is used by the equations 5.23 and 5.24 to calculate, at first, the specific offset for each subject during the initial condition where the eye angles are null, and then reading the P_x, P_y in the image plane all corresponding eye angles ($\tilde{\theta}_e$, $\tilde{\psi}_e$) are determined. Moreover, both data with and without VOR correction were collected and compared. The accuracy was calculated in terms of median and dispersion of the gaze error, where the error between the real observed position and the estimated value in pixels (e_pixel) is transformed into degrees according to the equation 5.25

$$e_{degree} = arctan(\frac{e_{pixel}}{d_{pixel}}) \tag{5.25}$$

where d_{pixel} represents the distance between the subject and the calibration plane. Tables 5.7 - 5.8 show median and dispersion per subject expressed in degrees in stationary head condition for the x and y axes. In particular, the first column refers to the values without any correction; the column "Screen point" refers to the error values in the image plane with the correction only actuated by IMU; and the column "Screen point + VOR" refers to the values in the image plane with the correction actuated by IMU and VOR together. The corrections are also reported for both two and three head rotation angles, respectively.

Tables 5.9 - 5.10 show median and standard deviation per subject expressed in degrees with free head movement condition for the x and y axes. The columns report the values as previous Tables 5.7, 5.8. The head movements are related to an average rotation amplitude of 20 degrees around the three axes.

The medians reported in Tables 5.9 - 5.10 are compared by using

Table 5.7: X accuracy with head in stationary mode

		3 angles		2 angles	
Subject	**Without Correction**	**Screen Point+VOR**	**Screen point**	**Screen Point+VOR**	**Screen point**
1	1.01±0.70	1.08±0.71	0.74±0.86	1.04±0.69	0.75±0.86
2	0.69±0.45	0.49±0.40	1.29±0.87	0.68±0.42	1.26±0.83
3	1.28±1.12	1.52±1.11	1.32±0.96	1.28±1.14	1.20±1.04
4	0.82±0.61	0.84±0.62	0.82±0.65	0.81±0.62	0.83±0.65
5	0.48±0.34	0.54±0.36	0.57±0.53	0.52±0.37	0.57±0.53
6	0.76±0.55	0.74±0.56	0.73±0.57	0.76±0.55	0.72±0.57
7	1.74±1.30	1.81±1.36	1.72±1.11	1.78±1.33	1.73±1.11
8	0.85±0.37	0.79±0.42	0.43±0.16	0.82±0.37	0.43±0.16
9	0.67±0.42	0.70±0.52	0.67±0.52	0.71±0.48	0.65±0.51
10	1.51±0.59	1.54±0.60	1.43±0.60	1.50±0.61	1.44±0.59
11	1.16±0.72	1.21±0.70	1.25±1.07	1.21±0.63	1.26±1.08
Median	**0.85±0.31**	**0.84±0.37**	**0.82±0.37**	**0.82±0.32**	**0.83±0.36**

Table 5.8: Y accuracy with head in stationary mode

		3 angles		2 angles	
Subject	**Without Correction**	**Screen Point+VOR**	**Screen point**	**Screen Point+VOR**	**Screen point**
1	1.93±1.01	1.93±1.14	1.98±1.22	2.06±1.01	1.95±1.23
2	1.38±0.97	2.26±1.38	1.29±1.08	1.75±1.14	1.50±1.08
3	1.62±1.22	1.31±0.97	1.85±1.50	1.64±1.24	1.75±1.34
4	0.84±0.82	0.97±0.90	1.29±1.06	0.98±0.88	1.29±1.06
5	1.37±0.52	1.32±0.46	1.37±0.56	1.34±0.49	1.36±0.55
6	3.20±1.41	3.22±1.41	3.07±1.44	3.21±1.43	3.07±1.44
7	2.82±1.38	2.83±1.37	2.84±1.39	2.82±1.39	2.85±1.38
8	0.89±0.52	0.77±0.54	0.98±0.54	0.88±0.53	0.92±0.55
9	1.65±1.36	1.90±1.08	2.47±0.80	1.95±1.14	2.46±0.79
10	0.92±0.60	0.85±0.60	0.79±0.55	0.86±0.60	0.80±0.54
11	1.70±0.92	1.61±1.05	1.40±0.96	1.61±0.95	1.42±0.92
Median	**1.62±0.54**	**1.61±0.64**	**1.40±0.62**	**1.64±0.56**	**1.50±0.60**

Table 5.9: X accuracy with the head free to move, corrections are reported with two and three angles

Subject	Screen Point+VOR		Screen point	
	3 Angles	2 Angles	3 Angles	2 Angles
1	1.74±1.26	1.64±1.22	11.49±15.35	11.51±15.39
2	0.85±0.52	0.64±0.37	5.32±4.54	5.31±4.54
3	1.15±1.10	1.08±1.23	5.15±4.28	5.13±4.32
4	0.56±0.37	0.50±0.36	3.10±1.93	3.10±1.95
5	0.37±0.24	0.35±0.21	3.57±2.40	3.63±2.50
6	0.69±0.14	0.24±0.22	9.84±5.09	9.76±4.97
7	0.65±0.30	0.69±0.39	5.67±1.51	5.63±1.54
8	0.80±0.84	0.68±0.78	11.89±5.76	11.95±5.80
9	1.74±1.18	2.14±1.00	4.68±1.82	4.65±1.82
10	1.74±1.02	1.87±0.96	1.58±0.85	1.55±0.86
11	1.39±1.10	1.45±1.04	7.73±6.27	7.68±6.33
Median	**0.85±0.44**	**0.69±0.55**	**5.32±2.81**	**5.31±2.81**

Table 5.10: Y accuracy with the head free to move, corrections are reported with two and three angles

Subject	Screen Point+VOR		Screen point	
	3 Angles	2 Angles	3 Angles	2 Angles
1	3.13±1.45	3.00±1.30	10.91±11.99	10.85±12.06
2	1.24±1.06	1.30±0.98	4.75±4.87	4.71±4.77
3	1.34±0.99	1.24±0.78	1.11±0.84	1.18±0.96
4	3.09±1.02	3.33±1.01	3.44±2.40	3.44±2.41
5	1.18±0.88	1.16±0.93	2.18±1.90	2.12±1.85
6	6.50±0.77	5.76±0.69	15.55±1.92	15.76±1.94
7	2.20±1.03	2.04±0.93	10.98±5.02	10.82±4.84
8	1.78±1.07	1.84±1.08	8.50±5.11	8.51±5.11
9	2.25±1.85	2.20±1.74	3.11±2.58	3.25±2.56
10	1.51±1.06	1.40±0.98	1.89±1.50	1.93±1.57
11	1.21±1.03	1.28±0.99	4.11±4.40	4.07±4.39
Median	**1.78±1.05**	**1.84±0.98**	**4.11±3.95**	**4.70±3.94**

Table 5.11: Kruskal-Wallis test between stationary mode and head movement with/without VOR correction

	With VOR	Without VOR
X Stationary mode	0.998	0.959
Y Stationary mode	0.940	0.957
X head movement	0.881	3.27E-005
Y head movement	0.563	1.5E-3

a dispersion analysis. Tables 5.11 show the Kruskal-Wallis test [133] in stationary head (stationary) and head free to move (movement) conditions, with and without VOR correction to verify that the median and dispersion of the groups were all equal. More specifically, the Kruskal-Wallis test return the probability that the different samples did not belong to the same population.The Figures 5.31 and 5.32 show the graphical analysis of the Kruskal-Wallis statistical test.

Table 5.12: Execution time

Task	**Time(ms)**
DCT	3.25
Beam Tracking+Fitting ellipse	22.51
Total software	34.47

Table 5.12 shows the execution time of the two mains tasks and the whole software integrated into the system. The work frequency is about 29 Hz, which is greater than the camera sample frequency

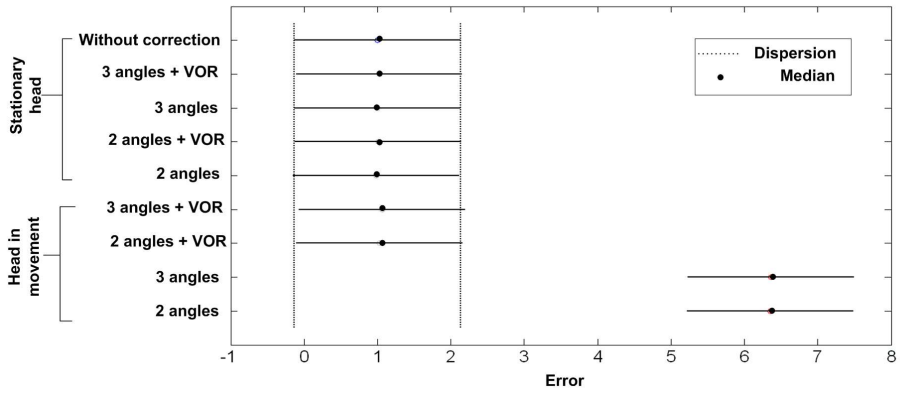

Figure 5.31: Graphical analysis of the statistical comparison of the median and dispersion along with the x axis,in stationary mode and with head movements with and without VOR correction

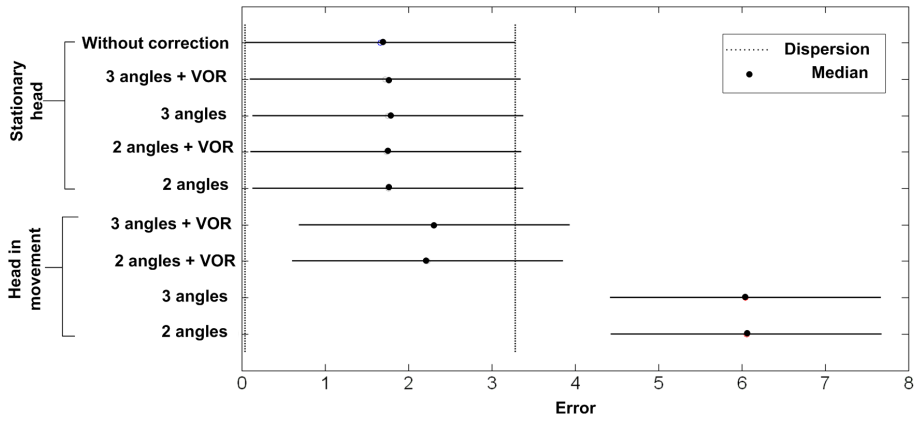

Figure 5.32: Graphical analysis of the statistical comparison of the median and dispersion along with the y axis, in stationary mode and with head movements with and without VOR correction

(25Hz), and therefore the real-time issue is respected.

5.8 Eye tracking and pupil size variation as response to affective stimuli

The associations between emotions and physiological reactions controlled by the ANS are complex, but anger, for example, has been associated with a higher heart rate than happiness; moreover, anger has been associated to a higher finger temperature than fear [135], [136]. It is interesting to investigate the relationship between emotions and information coming from the eyes, i.e. pupil size variation and eye tracking. In fact, it is known that pupil dilations and constrictions are governed by the ANS [137]. Previous studies have suggested that pupil size variation is also related to both cognitive and affective information processing. Nowadays, eye-tracking technology development (e.g. ease of use, improved accuracy, and enhanced sampling rate) offers the possibility for unobtrusive monitoring of emotion-related reactions, because no sensors need to be attached to the user. In order to be able to evaluate the possibilities of using pupil size measurement as well as pupil tracking for detecting emotional responses, it is necessary to understand how emotions and eye features relate to each other.

Concerning eye tracking, people can consider and discard various aspects of a task rather quickly (in less than 200 ms). Eye movements can provide detailed estimates of what information an individual is considering. Eye tracking is becoming an increasing popular measure of cognitive and affective information processing [138]. By gathering

data on the location and duration of eye fixations, many inferences about the cognition structure could be made. The use of eye tracking in estimating cognitive or affective states can be focused on two assumptions: the immediacy (people process information as it is seen) and the eye-mind (the eye remains fixated on an object while the object is being processed). A variety of eye-tracking methods exist. In terms of the data collected from the eye, two popular methods are: 1) shining a light on the eye and detecting corneal reflection and 2) simply taking visual images of the eye and then locating the dark pupil area. In the next section, a study that explore eye tracking and pupil area variation in response to stimulation using images from IAPS will be discussed. The aim of the study is to explore the association between eye information and emotional image categories. More specifically, by using HATCAM it is possible to acquire pupil variation together with eye gaze trajectory and fixation time during the exposition of subjects to affective images. The goal of the study was to identify characteristic features from pupil size variation and eye tracking by means of classification methods to distinguish between neutral and arousal-elicited states.

Ten subjects, nine males and one female, volunteered to participate in the experiment. None had ever suffered from mental pathologies and had average age of 26.8 with a standard deviation of 1.5. They were invited to perform the experiment in a room with a illumination condition achieved by white neon lighting equally distributed in the room with a power of 50 lumens. The subjects were asked to sit in a comfortable chair in front of a screen at a fixed distance of 70 cm. The HATCAM system was also equipped with a chin-support in order to avoid unwanted head movements. They were presented

with a sequence of images gathered from the IAPS database while wearing headsets for acoustical insulation. The slideshow was comprised of 5 sessions of images N, A, N, A, N, where N was a session of 5 neutral images, and A were sets of 5 images having a maximum level of arousal and the lowest valence, i.e. high negative affective impact. During the experimental test, all of the subjects were asked to look at the picture that appeared on the screen for 10 seconds. Each trial lasted about 25 minutes.

5.8.1 Post-processing and feature extraction

The post-processing phase is based on parameters extracted by Recurrence Quantification Analysis (RQA), along with fixation time and pupil area detection. Each image of the sequence was watched by the subjects for a time interval of 10 seconds. As the camera acquired with 25 fps, a grand total of 250 frames for each IAPS image was obtained. Each image can be represented as a matrix of $628x586$ pixels. The gaze point is mapped into a pixel in each frame. A matrix of 628 rows and 586 columns is constructed, where any position corresponding to the pixel coordinates associated to the gaze point is set to 1. In order to minimize errors due to the eye blinking, only the pixels retained for at least five frames (0.2 sec) are set to 1.

Recurrence Quantification Analysis

RQA [139] is a nonlinear analysis method which quantifies the number and duration of recurrences of a dynamical system represented by state space trajectories. Let RP be the Recurrence Plot matrix. Let $L_{ij}(x, y)$ be the set of eye-gaze points from the frame i of the

image j, where x and y are spatial coordinates. We applied the RQA to the set $G_m(x, y)$ defined as:

$$G_m(x, y) = \bigcup_i L_{ij}(x, y)$$

Therefore, for each image the following features are calculated:

Recurrence Rate (RR) is the percentage of recurrence points in an RP and corresponds to the correlation sum:

$$RR = \frac{1}{N^2} \sum_{x,y=1}^{N} G_m(x, y)$$

where N is the number of points on the phase space trajectory.

Determinism (DET) is the percentage of recurrence points which form diagonal lines:

$$DET = \frac{\sum_{l=lmin}^{N} lP(l)}{\sum_{i,j=1}^{N} R_{i,j}} \tag{5.26}$$

where $P(l)$ is the histogram of the lengths l of the diagonal lines.

Trapping Time (TT) is the average length of the vertical lines:

$$TT = \frac{\sum_{v=v_{min}}^{N} vP(v)}{\sum_{v=v_{min}}^{N} P(v)} \tag{5.27}$$

Averaged diagonal line length (L) is the average length of the diagonal lines:

$$L = \frac{\sum_{l=lmin}^{N} lP(l)}{\sum_{l=lmin}^{N} P(l)} \tag{5.28}$$

Entropy ($ENTR$) is the Shannon entropy of the probability distribution of the diagonal line lengths $p(l)$:

$$ENTR = -\sum_{l=lmin}^{N} p(l) \ln p(l) \tag{5.29}$$

Longest diagonal line (L_{max}) The length of the longest diagonal line:

$$L_{max} = max\left(\{l_i; i = 1, ..., N_l\}\right) \tag{5.30}$$

where N_l is the number of diagonal lines in the recurrence plot.

Fixation time

While watching each image, the subject's eye can be caught in specific detail. The fixation time of each pixel is defined as the absolute time during which the subject dwells on that pixel. A statistical distribution of fixation time over the fixed pixels is obtained, whose mode, which is defined here as $Tmax$, is used as an additional feature. $Tmax$ is calculated for each image and each subject during both neutral and arousal elicitation as:

$$Tmax = \max_{i=0}^{N}(t(P_i)) \tag{5.31}$$

where N is the number of points of gaze in the image, P_i is the i_{th} point of gaze, $t(P_i)$ is the fixation time of the i_{th} point of gaze, respectively.

Pupil Area Detection

The pupil area was approximated as an ellipse whose area is calculated for the pupillometry. To increase the robustness of the algorithm, an averaged area of both eyes is considered:

$$A_p = \frac{\pi r_a^l r_b^l + \pi r_a^r r_b^r}{2} \tag{5.32}$$

where A_p is the pupil area, r_a^l and r_b^l are the ellipse semi-axes of the left eye, r_a^r and r_b^r are the ellipse semi-axes of the right eye.

5.8.2 Classification

This aims is to classify two classes of images, labeled as neutral and arousal. Pattern recognition was performed by means of non-parametric K-Nearest Neighbor (K-NN) algorithm [140]. It is based on the "proximity" concept, i.e. an object is supposed to belong to the closest class in the n-dimensional feature space. After the K-NN training process, the performance of the classification task is commonly evaluated using the confusion matrix [141]. The generic element r_{ij} of the confusion matrix indicates how many times a pattern belonging to the class i was classified as belonging to the class j as a percentage. The more the confusion matrix is diagonal, the better

the classification. The matrix has to be read by columns. The training phase is carried out on 80% of the feature dataset while the testing phase on the remaining 20%. 40-fold cross-validation steps are performed in order to obtain unbiased classification results, allowing us to consider as Gaussian the classification result distributions, which can be therefore described as mean and standard deviations among the obtained 40 confusion matrices. The implementation was performed according to the following steps:

1. In the training phase, the K-NN algorithm just stores the training feature set together with the labels.

2. In the test phase, the K-NN algorithm calculates the n Euclidean distances between the new feature set and the n features of the whole training feature set as follows:

$$D_i = \sqrt{(p_{i1} - q_1)^2 + ... + (p_{in} - q_n)^2} = \sqrt{\sum_{j=1}^{n} (p_{ij} - q_j)^2} \qquad (5.33)$$

   where $P_i = (p_{i1}, p_{i2}, ..., p_{in})$ is i-th training feature set, $Q = (q_1, q_2, ..., q_n)$ is the new feature test set, n is the number of features.

Afterwards, the K-NN algorithm finds the K training feature sets that have the minimum distance from the new feature set. Among these, m training feature sets belong to the neutral class and $K - m$ belong to the arousal class. The new feature set is supposed to belong to the neutral class if $m > (K - m)$, i.e. $m > \frac{K}{2}$, otherwise to the arousal class.

Figure 5.33: Example of the points of gaze detected during a neutral elicitation. Gaze points are marked in red.

5.8.3 Classification Results

Figures 5.33 and 5.34 show the map of gaze points over two sample images, having neutral and the highest value of arousal, respectively. Gaze points are marked in red. At a glance, the neutral image (fig. 5.33) shows a more sparse spatial distribution of the gaze points than the arousal image, in which gaze points are mainly concentrated in confined areas. RQA analysis was used to quantify the gaze point distribution for each frame of each image. All of the extracted features from RQA and pupil area variation are not normally distributed, as confirmed by the Lilliefors test, [132], which returns a $p - value$ ($p<0.05$) rejecting the null hypothesis of normality. Accordingly, we used the Kruskal-Wallis test, [133], which is a non-parametric, one-way analysis of variance by ranks for testing the equality of population medians. Kruskal-Wallis is performed on ranked data, so the measurement observations are converted to their

Figure 5.34: Example of the points of gaze detected during a arousal elicitation. Gaze points are marked in red.

ranks in the overall data set. This test assumes an identically-shaped and scaled distribution for each group, except for any difference in medians. The null hypothesis is stated as the probability that the samples come from identical populations, regardless their distributions. In place of the mean of distributions we considered the median as a measure of location [134]. Having only two sets of features, one for arousal and one for neutral, Kruskal-Wallis test returned the probability that the two samples did not belong to the same population, or if there was a statistical difference between the two samples. Median and Median absolute deviations of all the RQA features are reported in Table 5.13. Statistical differences between neutral and arousal elicitation was found ($*p<0.01$ and $**p<0.001$).

In the table 5.14, the confusion matrix obtained from K-NN classifier after 40 fold-cross-validation steps is shown.

Table 5.13: Features extracted from RQA

Features	Neutral	Arousal
RR∗	0.0018±0.0002	0.0019±0.0002
DET∗∗	0.7311±0.0782	0.6373±0.0798
TT∗∗	2.5862±0.9024	2.0345±0.4368
L∗∗	2.9184±0.4564	2.6513±0.3181
ENTR∗∗	1.2592±0.2306	1.0501±0.2041
Lmax ∗∗	5.0000±1.2800	5.0000±1.0583
Tmax ∗∗	1.9600±0.3509	1.5600 ±0.3501
Pupil	195.39±19.689	197.27±17.106

Table 5.14: Confusion Matrix of K-NN classifier by using RQA and pupil size features

	Neutral	Arousal
Neutral	**90.2273±5.9622**	20.1136±10.8881
Arousal	9.7727±5.9622	**79.8864±10.8881**

5.9 Results and Discussions

In this chapter a new wearable eye tracking system was proposed. The system provides a technical innovation in both hardware and software. In order to increase the robustness of the device to the illumination change, seven photonormalization techniques have been investigated. The key approach was to evaluate, in terms of accuracy of eye tracking, pupillometry and execution time, seven photonormalization techniques under three different brightness conditions: Laboratory, Sunlight, and Darkness. The results showed that the DCT method is the most effective in terms of accuracy of gaze point and pupillometry. The NH is the best in terms of eye tracking but is less effective in the pupillometry; nevertheless, it results to be the fastest. The execution time is a crucial issue for the real-time processing according to the requirements introduced.
A new algorithm to compensate the head movement during the calibration process has been introduced. This algorithm acts to correct the calibration point on the image plane and the eye position. In order to reposition the eye, the VOR curves have been studied with an ad-hoc experiment. The efficacy of the correction with head movements has been proven by statistical analysis taking into account the median of error expressed in degrees and the dispersion. This is an important result because the system is able to overcome a, important problem of many commercial EGT systems: the users are relieved from keeping their head in stationary mode when the correction is used. The estimated median error with head movement and correction is approximately 0.85° on x direction and 1.78° on x direction with dispersion, respectively 0.44◦ and 1.05◦.

The obtained accuracy results confirm the reliability and robustness of the proposed system also under different illumination conditions thanks to a photonormalization algorithm.
Therefore, the proposed EGT showed reliable and accurate global results. The large dispersion is produced by the impossibility of the subjects to continuously fix the target (the gaze point oscillates around the target). Moreover, it must be noted that there is a substantial difference between the accuracy along the X-axis and along the Y-axis which comes from two main factors. The first source is related to the angular position of the camera (above of the eyes), which reduces the accuracy of the vertical positions of the pupil. The second refers to the eye features tracking: if the gaze is directed towards the very lowest part, the tracking of the pupil is occluded by the eyelids. Moveover, the partial occlusion of the highest part of the FOV (the subject is not able to fix the point) by the mirror during the VOR vertical calibration, most likely reduced the accuracy of the system. Table 5.12 shows the execution time of the software's main task which shows that the EGT can be used in real time. However, in the best of cases, the system cannot record fast saccadic movements (which are about 30 msec) but can detect slow saccadic movements (100 to 120 msec).
The proposed system makes use of simple, low cost and commercially available hardware. The fully ecological approach by designing a wearable device which is extremely light, unobtrusive, attractive and aesthetic have provided a good acceptability by the end users, except for the fact that some subjects complaint about a partial occlusion of the very highest part of the FOV by the mirror, specifically with the head band solution. It is possible to resolve this problem by

modifying the arm in length and shape. In opposition the users liked the absence of wires that allowed for free movements. The main limit of the system is the camera, which has a low resolution and frame rate. Therefore further efforts will be devoted to integrating a low cost, high-speed camera in order to capture fast saccadic movements are about 30 msec. Moreover, the use of a camera with a high resolution could provide better accuracy than actual accuracy. However a good tradeoff between the performance the system in terms of work frequency and accuracy must be found.
The optimization of the system will be addressed to develop new multithreading algorithms based on the interocular distance in order to obtain a 3D eye tracking system.
The investigation on eye tracking and pupil size variation, in response to emotional elicitation induced by IAPS images to distinguish between neutral to arousal states, has been reached by the introduction of a novel methodology by means of features extracted from RQA. This choice has been motivated by the analogy between the bi-dimensional image containing eye-gaze points and the matrix commonly used for Recurrence Plot [142]. Both, indeed, are matrices of zeroes and ones. In addition to the features from RQA, the elliptic area of the pupil and fixation time were included in the feature space for arousal and neutral classification. Pattern recognition was performed by means of K-Nearest Neighbor (K-NN) classifier. After the K-NN training process, the performance of the classification task was evaluated by using the confusion matrix. It was randomized for 40-fold cross-validation steps to avoid bias. Results are reported in Table 5.14. As it can be seen results are very satisfactory. The percentage of successful recognition is about 90% for neutral images

and about 80% for images at high arousal. It means that eye gaze, both in terms of pupil tracking and size, can be a viable means to discriminate different affective states. Future work will progress to exploit eye gaze together with peripheral physiological signals in the field of mental care or to support the patient emotion recognition in the FACET platform.

Chapter 6

Conclusions and Future Works

In this chapter the main outcomes of this research work will be summarized, and I will outline future research directions, stressing the multiple application fields of the designed devices. This work has given an overview of the major assistive technologies for people affected by autistic spectrum disorders. Research suggests that people with autism generally feel comfortable in predictable environments, and more specifically, enjoy interacting with computer and robots. Based on this evidence, this PhD dissertation has been designed to develop a complex platform named FACET that is able to adapt a therapy using a robot endowed with the ability to sense, adapt and respond to a patient's emotional and mental states, enabling autistic subjects to learn empathy and gradually enhance their social competence. Therefore a structured therapeutic environment is used in which the subject's behavior and responses are monitored using a

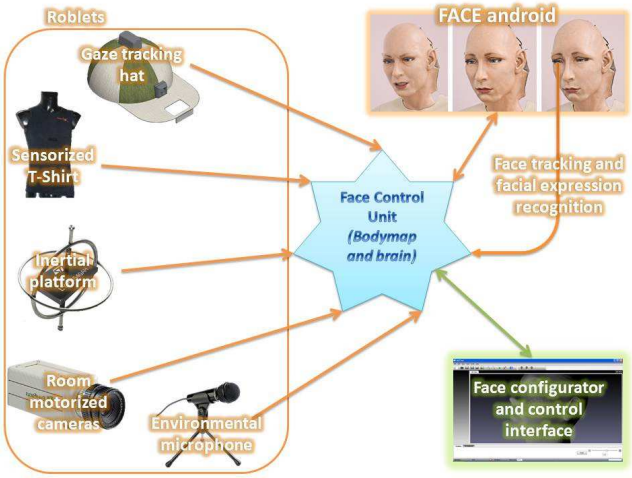

Figure 6.1: FACET platform

multiplicity of sensors that are then processed and fed back to the android to modulate and modify its expressions. Physiological signals (HR, HRV, RR, EMG and EDA) and eye gaze are acquired using a comfortable, unobtrusive ,sensorized shirt and a cap containing integrated cameras and mini gyroscopes. The design and development of the android cognitive architecture, inspired to imitate the brain functions, has been proposed on the Field Programmable Gate Array (FPGA) in order to process the sensor signals and to modulate its behavior. In fact, the real-time interaction with the patient necessitates a system having the following requirements:

- Computational Complexity
- Speed performance

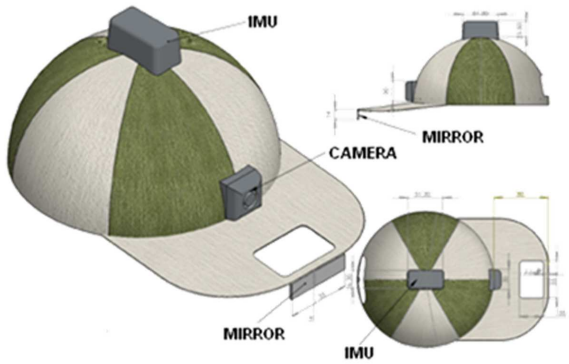

Figure 6.2: Version I: "Baseball-like" hat configuration.

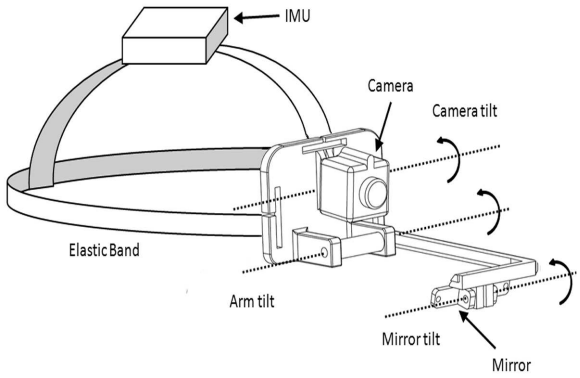

Figure 6.3: Version II: "Head band" configuration.

- Low size
- Low power consumption

The best architecture that provides a good compromise between all requirements is the implementation of the ANN on FPGA.
The most complex task in implementing the hardware ANNs is the nonlinear activation function. A *new mathematical approach* inspired by the "area method" originally proposed by Francesco Paolo Cantelli in 1905 [73] has been introduced to approximate the sigmoid and hyperbolic tangent functions [85]. The area method adopted has provided a very simple methodology to approximate the sigmoid function, hyperbolic tangent and their derivative. Moreover, the obtained results have the least absolute and relative errors. In fact, as reported in Table 4.2 referring to the approximation of the sigmoid, *the area method has provided the least absolute mean error and the least absolute maximum error*. It has the least average relative error. The relative maximum error is comparable with the results of Bajger [77].
The PWL approximation of the hyperbolic tangent, extrapolated by the mathematical relationship between this function and the sigmoid (Table 4.4), has the best results in terms of average and maximum absolute error. Similar considerations can be made for the derivative of the hyperbolic tangent, extrapolated by the mathematical relationship between this function and the sigmoid derivative.
The PWL approximations of the hyperbolic tangent and its derivative, directly obtained by the area method, are comparable with the results obtained by the mathematical relationship with sigmoid.
The sigmoid digital implementation has been compared with the

state of the art. Scheme 1 shows in Figure 4.20 the best performance with respect to the other implementation, except for Tommiska [64] which is comparable in terms of error. However, the work [64] reports a very low number of precision bits in input. To obtain adequate convergence of a network, the study of Holt [70] has been considered as well as the recent study of Savich [72]. Therefore, based on the convergence study, more bits precision of the state of the art has been considered to the detriment of the performance of gate complexity (Fig.4.21). A flexible topology (Virtual Reconfigurable Network) inspired by an evolving hardware architecture has been proposed. The architecture has been tested on the logical-XOR problem reporting the capability to reconfigure the architecture of the network and to learn new behavior. In fact, the ad hoc Genetic Algorithm (GA) has guaranteed the evolution of different topologies. Future work could address new hardware architecture based on Network-on-a-Chip (NOC), which could be an emergent approach to design a neural communication subsystems with a high level of parallelism and flexibility.

A new wearable eye tracking system has been proposed. The system has reported new technical innovation in both hardware and software [brevetto]. *An important novelty has been introduced adopting photonormalization techniques to increase the robustness of the device duringthe illumination change.* The key approach has been to evaluate, in terms of accuracy of eye tracking, pupillometry and execution time, seven photonormalization techniques under three different brightness conditions: Laboratory, Sunlight, Darkness. The results have showed that the DCT method is the most effective in terms of accuracy of gaze point and pupillometry.

Figure 6.4: A control subject wear HATCAM.

A *new algorithm to compensate head movement* during the calibration process has been introduced. This algorithm acts to correct the calibration point on the image plane and eye position. In order to reposition the eyes, the Vestibulo-Ocular Reflex (VOR) is involved which is responsible for the compensatory eye rotations in response to head motion, as sensed by the vestibular organs in the inner ear [131]. More specifically, during head rotations the VOR rotates the eyes in the opposite direction to maintain a stable gaze [114]. The VOR curves have been calculated by an ad hoc experiment. During the experiment, subjects were asked to rotate their head about the horizontal and vertical axes keeping a fixed gaze on the point placed in front of them. In the latter case, head rotation is equal but opposite to the eye rotations. Moreover, by taking into account the acquired head rotations, it has been possible to correct the projection of the calibration point in the exact position on the image plane

(neglecting the translations). *The efficacy of correction in head movements has been proven by statistical analysis by taking into account the median error expressed in degrees and the dispersion. This is an important result because the system is able to overcome an important problem of many commercial EGT systems: the users do not have to keep their head stationary when the correction is used.* The estimated accuracy in head movement and correction is approximately 0.85° in the x-direction and 1.78° in the y direction with a dispersion, respectively, of 0.44° and 1.05° in the y-direction.

Therefore, the proposed EGT shows reliable and accurate global results. The large dispersion is produced by the inability of the subjects to continuously fixate on the target (the gaze point oscillates around the target). However, a note must be made about the substantial difference between the accuracy along the X and Y axes that comes from two main factors. The first is related to the angular position of the camera (above of the eyes), which reduces the accuracy of the vertical positions of the pupil. The second relates to the eye tracking; if the gaze is directed downward , pupil tracking is obstructed by the eyelids. Moveover, the partial obstruction of the upward portion of the FOV (the subject is not able to fixate on the point) by a mirror during VOR vertical calibration probably reduced the accuracy of the system.

The execution time of the software's main task demonstrates that HATCAM can be used in real time. However, in a best case scenario, the system cannot record fast saccadic movements (which range around 30 msec), but can detect slow saccadic movements (from 100 to 120 msec).

The proposed system makes use of simple, low cost and commercially-

available hardware. *The ecological approach, by designing a wearable device which is extremely light, unobtrusive, attractive and aesthetically pleasing ,has provided a good acceptability by the end users,* (Fig.6.4) although some subjects complained about a partial obstruction of the upward portion of the FOV by a mirror, specifically with regards to the head band solution. It will be possible to resolve this problem by modifying the arm in length and shape. Other users liked the absence of wires allowing for free movement.
The main limit of the system is the camera, which has a low resolution and frame rate. Therefore, further efforts will be devoted to integrating a low cost, high-speed camera in order to capture fast saccadic movements that range around 30 msec. Moreover, the use of a camera with a high resolution could provide a better accuracy than actual accuracy. However, a good tradeoff needs to be found between the system performance in terms of work frequency and accuracy.
The optimization of the system will be addressed to develop a new, multi-threading algorithm based on the interocular distance in order to obtain a 3D eye-tracking system.
The investigation on eye tracking and pupil size variation in response to emotional reaction induced by IAPS images (to distinguish between neutral and arousal states)has been carried out by introducing a novel methodology through feature extraction from RQA. This choice has been motivated by the analogy between the bi-dimensional image containing eye-gaze points and the matrix commonly used for Recurrence Plot [142]. Both, indeed, are matrices of zero and ones. In addition to the features from RQA, the elliptic area of the pupil and the fixation time were included in the feature space

for arousal and neutral classification. Pattern recognition was performed with K-Nearest Neighbor (K-NN) classifier. After the K-NN training process, the performance of the classification task was evaluated by using the confusion matrix. It was randomized for 40-fold cross-validation steps to avoid bias. Results are reported in Table 5.14, and, as it can be seen, the results are very satisfactory. The percentage of successful recognition is about 90% for neutral images and about 80% for images at high arousal. This means that eye gaze, both in terms of pupil tracking and size, can be a viable means to discriminate different affective states.

Future work will progress to exploit eye gaze together with peripheral physiological signals in the field of mental care or to support the patient emotion recognition in the FACET platform.

In conclusion, HATCAM has the strong potential to investigate attention disorders early in autistic infants during therapy. Moreover, it can used to test and validate the appearance of the android in order to increase attention and the interaction of a young patient. The possibility of obtaining the gaze point in real-time enables the therapist to modify his actions according to the locus of the patient's attention. These are not the only applications. In fact, autism is usually diagnosed by 3 years of age. The diagnosis is purely clinical and there are no laboratory tests that confirm or disprove the diagnosis. The development of HATCAM, which is a user-friendly technology suitable of observing very young people in familiar environments, could be used to diagnose autism early and to assess the therapy of the neurodevelopment disorder.

Bibliography

[1] E. Bleuler and G. Aschaffenburg. *Dementia praecox: oder Gruppe der Schizophrenien*. F. Deuticke, 1911.

[2] L. Kanner et al. Autistic disturbances of affective contact. *Nervous child*, 2(2):217–230, 1943.

[3] H. Asperger. Die autistichen psychopathen im kindesalter [Autistic psychopathy in infancy]. *Archiv f ”ur Psychiatrie und Nervenkrankheiten [Archive for Psychiatry and Nerve Disease]*, 117:76–136, 1944.

[4] B. Bettelheim. *The empty fortress: Infantile autism and the birth of the self*. Free Pr, 172.

[5] American Psychiatric Association. Committee on Nomenclature, Statistics, and American Psychiatric Association. Committee on Statistics. *Diagnostic and statistical manual of mental disorders*, volume 2. American Psychiatric Association Washington, DC, 1968.

[6] E.A. Tinbergen and N. Tinbergen. *Early childhood autism: An ethological approach*. Parey, 1972.

[7] N. Tinbergen, E.A. Tinbergen, M. Welch, and M. Zappella. *'Autistic'Children: New Hope for a Cure*. Allen & Unwin London, 1983.

[8] S. Wolff. Autism: Explaining the enigma. *Journal of Child Psychology and Psychiatry*, 45(1):172–173, 2004.

[9] L. Wing and A. Attwood. Syndromes of autism and atypical development. *Handbook of autism and pervasive developmental disorders*, 1:3–19, 1987.

[10] American Psychiatric Association. Committee on Nomenclature, Statistics, and American Psychiatric Association. Committee on Statistics. *Diagnostic and statistical manual of mental disorders*, volume 3. American Psychiatric Association Washington, DC, 1980.

[11] World Health Organization. *Mental disorders: A glossary and guide to their classification in accordance with the 10th Revision of the Internazional Classification of Diseases*. 1987.

[12] American Psychiatric Association. Committee on Nomenclature, Statistics, and American Psychiatric Association. Committee on Statistics. *Diagnostic and statistical manual of mental disorders*, volume 4. American Psychiatric Association Washington, DC, 1994.

[13] G. Rizzolati, L. Fadiga, V. Gallese, and L. Fogassi. Premotor cortex and the recognition of motor actions. *Cognitive brain research*, 3(2):131–141, 1996.

BIBLIOGRAPHY

[14] L. Fadiga, L. Fogassi, G. Pavesi, and G. Rizzolatti. Motor facilitation during action observation: a magnetic stimulation study. *Journal of neurophysiology*, 73(6):2608, 1995.

[15] J. Decety, J. Grezes, N. Costes, D. Perani, M. Jeannerod, E. Procyk, F. Grassi, and F. Fazio. Brain activity during observation of actions. Influence of action content and subject's strategy. *Brain*, 120(10):1763, 1997.

[16] M. Iacoboni, R.P. Woods, M. Brass, H. Bekkering, J.C. Mazziotta, and G. Rizzolatti. Cortical mechanisms of human imitation. *Science*, 286(5449):2526, 1999.

[17] JA Pineda, D. Brang, E. Hecht, L. Edwards, S. Carey, M. Bacon, C. Futagaki, D. Suk, J. Tom, C. Birnbaum, et al. Positive behavioral and electrophysiological changes following neurofeedback training in children with autism. *Research in Autism Spectrum Disorders*, 2(3):557–581, 2008.

[18] Pascual-Leone A. Oberman L.M., Maria-Tormos J. Preliminary Evidence for Dysfunctional Cortical Plasticity Mechanisms in Autism Spectrum Disorders. Number 38. Washington D.C., 2008.

[19] Richard et al. Anney. A genome-wide scan for common alleles affecting risk for autism. *Human Molecular Genetics*, 19(20):4072–4082, 2010.

[20] J. Scholtz. Theory and evaluation of human robot interactions. 2003.

[21] J. Han, J. Lee, and Y. Cho. Evolutionary role model and basic emotions of service robots originated from computers. In *Robot and Human Interactive Communication, 2005. ROMAN 2005. IEEE International Workshop on*, pages 30–35. IEEE, 2005.

[22] K. Dautenhahn. Robots as social actors: Aurora and the case of autism. In *Proc. CT99, The Third International Cognitive Technology Conference, August, San Francisco*, pages 359–374, 1999.

[23] B. Robins, K. Dautenhahn, R. Boekhorst, and A. Billard. Robotic assistants in therapy and education of children with autism: Can a small humanoid robot help encourage social interaction skills? *Universal Access in the Information Society*, 4(2):105–120, 2005.

[24] R. el Kaliouby and P. Robinson. THERAPUETIC VERSUS PROSTHETIC ASSISTIVE TECHNOLOGIES: THE CASE OF AUTISM.

[25] K.M. Colby. The rationale for computer-based treatment of language difficulties in nonspeaking autistic children. *Journal of Autism and Developmental Disorders*, 3(3):254–260, 1973.

[26] S. Powell. The use of computers in teaching people with autism Autism on the agenda: papers from the National Autistic Society conference 1996. *National Autistic SocietyLondon, UK*.

[27] D. Moore. Computers and people with autism. *Asperger Syndrome*, pages 20–21, 1998.

BIBLIOGRAPHY

[28] D. Murray. Autism and information technology: therapy with computers. *Autism and learning: a guide to good practice*, pages 100–117, 1997.

[29] D. Moore, P. McGrath, and J. Thorpe. Computer-aided learning for people with autism–a framework for research and development. *Innovations in Education and Teaching International*, 37(3):218–228, 2000.

[30] Fun with feelings. http://www.digisoftdirect.com/products/funwithfeelings.html. 2004.

[31] G. Rajendran and P. Mitchell. Computer mediated interaction in Asperger's syndrome: the Bubble Dialogue program. *Computers & Education*, 35(3):189–207, 2000.

[32] A. Jones and C. Selby. The use of computers for self-expression and communication. *Journal of Computing in Childhood Education*, 1997.

[33] B. Galitsky. Using mental simulator for emotional rehabilitation of autistic patients. *St. Augustine, FL*, 2003.

[34] O. Golan and S. Baron-Cohen. Systemizing empathy: Teaching adults with Asperger syndrome or high-functioning autism to recognize complex emotions using interactive multimedia. *Development and psychopathology*, 18(02):591–617, 2006.

[35] S. Baron-Cohen, O. Golan, S. Wheelwright, and J.J. Hill. Mind reading: The interactive guide to emotions. *London: Jessica Kingsley Limited*, 2004.

[36] JJ Cromby, PJ Standen, J. Newman, and H. Tasker. Successful transfer to the real world of skills practised in a virtual environment by students with severe learning difficulties. In *Proceedings of the European conference on disability, virtual reality and associated technology*, pages 103–107, 1996.

[37] D. Strickland. Virtual reality for the treatment of autism. *Virtual Reality in Neuro-Psycho-Physiology: Cognitive, Clinical, and Methodological Issues in Assessment and Rehabilitation*, pages 81–86, 1997.

[38] L. Cheng, G. Kimberly, and F. Orlich. Kidtalk: Online therapy for aspergers syndrome. *Microsoft Research*, 2002.

[39] H. Kozima and J. Zlatev. An epigenetic approach to human-robot communication. In *Robot and Human Interactive Communication, 2000. RO-MAN 2000. Proceedings. 9th IEEE International Workshop on*, pages 346–351. IEEE, 2002.

[40] M. Lungarella, G. Metta, R. Pfeifer, and G. Sandini. Developmental robotics: a survey. *Connection Science*, 15(4):151–190, 2003.

[41] K.F. MacDorman and H. Ishiguro. The uncanny advantage of using androids in cognitive and social science research. *Interaction Studies*, 7(3):297–337, 2006.

[42] B. Robins, K. Dautenhahn, and P. Dickerson. From isolation to communication: a case study evaluation of robot assisted play for children with autism with a minimally expressive humanoid robot. In *Advances in Computer-Human Interactions,*

2009. ACHI'09. Second International Conferences on, pages 205–211. IEEE, 2009.

[43] M. Blow, K. Dautenhahn, A. Appleby, C.L. Nehaniv, and D. Lee. The art of designing robot faces: Dimensions for human-robot interaction. In *Proceedings of the 1st ACM SIGCHI/SIGART conference on Human-robot interaction*, pages 331–332. ACM, 2006.

[44] M. Blow, K. Dautenhahn, A. Appleby, C.L. Nehaniv, and D.C. Lee. Perception of robot smiles and dimensions for human-robot interaction design. In *Robot and Human Interactive Communication, 2006. ROMAN 2006. The 15th IEEE International Symposium on*, pages 469–474. IEEE, 2007.

[45] A. Billard, B. Robins, K. Dautenhahn, and J. Nadel. Building robota, a mini-humanoid robot for the rehabilitation of children with autism. *RESNA Assistive Technology Journal*, 19(1):37–49, 2006.

[46] C. Plaisant, A. Druin, C. Lathan, K. Dakhane, K. Edwards, J.M. Vice, and J. Montemayor. A storytelling robot for pediatric rehabilitation. In *Proceedings of the fourth international ACM conference on Assistive technologies*, pages 50–55. ACM, 2000.

[47] H. Kozima and H. Yano. A robot that learns to communicate with human caregivers. In *Proceedings of the First International Workshop on Epigenetic Robotics*, pages 47–52. Citeseer, 2001.

[48] H. Kozima, C. Nakagawa, and H. Yano. Using robots for the study of human social development. In *AAAI Spring Symposium on Developmental Robotics*, pages 111–114, 2005.

[49] H. Kozima. A Babybot that Explores the Social Environment. *Socially intelligent agents: creating relationships with computers and robots*, page 157, 2002.

[50] H. Kozima, C. Nakagawa, and H. Yano. Attention coupling as a prerequisite for social interaction. In *Robot and Human Interactive Communication, 2003. Proceedings. ROMAN 2003. The 12th IEEE International Workshop on*, pages 109–114. IEEE, 2003.

[51] T. Shibata, T. Mitsui, K. Wada, A. Touda, T. Kumasaka, K. Tagami, and K. Tanie. Mental commit robot and its application to therapy of children. In *Advanced Intelligent Mechatronics, 2001. Proceedings. 2001 IEEE/ASME International Conference on*, volume 2, pages 1053–1058. IEEE, 2002.

[52] T. Shibata, K. Wada, T. Saito, and K. Tanie. Robot assisted activity for senior people at day service center. In *Proc. of Intl Conf. on Information Technology in Mechatronics*, pages 71–76, 2001.

[53] A. Mihailidis and E.F. LoPresti. Cognitive Assistive Technology.

[54] K.M. Dalton, B.M. Nacewicz, T. Johnstone, H.S. Schaefer, M.A. Gernsbacher, HH Goldsmith, A.L. Alexander, and R.J.

Davidson. Gaze fixation and the neural circuitry of face processing in autism. *Nature Neuroscience*, 8(4):519–526, 2005.

[55] G. Vivanti, A. Nadig, S. Ozonoff, and S.J. Rogers. What do children with autism attend to during imitation tasks? *Journal of experimental child psychology*, 101(3):186–205, 2008.

[56] A. Klin, W. Jones, R. Schultz, and F. Volkmar. The enactive mind, or from actions to cognition: lessons from autism. *Philosophical Transactions B*, 358(1430):345, 2003.

[57] L. Piccardi, B. Noris, G. Schiavone, F. Keller, C. Von Hofsten, and AG Billard. Wearcam: A head mounted wireless camera for monitoring gaze attention and for the diagnosis of developmental disorders in young children. In *Robot and Human interactive Communication, 2007. RO-MAN 2007. The 16th IEEE International Symposium on*, pages 594–598. Citeseer, 2008.

[58] G. Pioggia, R. Igliozzi, M.L. Sica, M. Ferro, F. Muratori, A. Ahluwalia, and D. De Rossi. Exploring emotional and imitational android-based interactions in autistic spectrum disorders. *JCR*, 1(1):49, 2008.

[59] D. Mazzei, L. Billeci, A. Armato, N. Lazzeri, A. Cisternino, G. Pioggia, R. Igliozzi, F. Muratori, A. Ahluwalia, and D. De Rossi. The FACE of autism. In *RO-MAN, 2010 IEEE*, pages 791–796. IEEE.

[60] N. Mavridis and D. Hanson. The IbnSina Center: An augmented reality theater with intelligent robotic and virtual char-

acters. In *Robot and Human Interactive Communication, 2009. RO-MAN 2009. The 18th IEEE International Symposium on*, pages 681–686. IEEE, 2009.

[61]

[62] C. Eppner, J. Sturm, M. Bennewitz, C. Stachniss, and W. Burgard. Imitation learning with generalized task descriptions. In *Robotics and Automation, 2009. ICRA'09. IEEE International Conference on*, pages 3968–3974. IEEE, 2009.

[63] D. Floreano and C. Mattiussi. *Bio-inspired artificial intelligence: theories, methods, and technologies*. 2008.

[64] MT Tommiska. Efficient digital implementation of the sigmoid function for reprogrammable logic. In *Computers and Digital Techniques, IEE Proceedings-*, volume 150, pages 403–411. IET, 2003.

[65] S. Sahin, Y. Becerikli, and S. Yazici. Neural Network Implementation in Hardware Using FPGAs. In *Neural Information Processing*, pages 1105–1112. Springer, 2006.

[66] J. Zhu, GJ Milne, and BK Gunther. Towards an FPGA based reconfigurable computing environment for neural network implementations. In *Artificial Neural Networks, 1999. ICANN 99. Ninth International Conference on (Conf. Publ. No. 470)*, volume 2, pages 661–666. IET, 2002.

[67] K. Basterretxea, JM Tarela, and I. Del Campo. Digital design of sigmoid approximator for artificial neural networks. *Electronics Letters*, 38(1):35–37, 2002.

[68] P. Moerland and E. Fiesler. Neural network adaptations to hardware implementations. *Handbook of neural computation*, page 2, 1997.

[69] S. Draghici. On the capabilities of neural networks using limited precision weights. *Neural Networks*, 15(3):395–414, 2002.

[70] JL Holi and J.N. Hwang. Finite precision error analysis of neural network hardware implementations. *Computers, IEEE Transactions on*, 42(3):281–290, 2002.

[71] E. Gustafsson. A mapping of a feedback neural network onto a SIMD architecture. *Research Report CDv-8901, Centre for Computer Science, Halmstad University*, 1989.

[72] A.W. Savich, M. Moussa, and S. Areibi. The impact of arithmetic representation on implementing MLP-BP on FPGAs: A study. *Neural Networks, IEEE Transactions on*, 18(1):240–252, 2006.

[73] FP Cantelli. Sull'adattamento delle curve ad una serie di misure o di osservazioni, 1905.

[74] P.F. Verhulst. Recherches mathématiques sur la loi d'accroissement de la population. 1845.

[75] D.J. Myers and R.A. Hutchnson. Efficient implementation of piecewise linear activation for digital VLSI neural network. *Electronics Letters*, 25(24):1662–1663, 1989.

[76] S. Vassiliadis, M. Zhang, and J.G. Delgado-Frias. Elementary function generators for neural-network emulators. *Neural Networks, IEEE Transactions on*, 11(6):1438–1449, 2002.

[77] M. Bajger and A. Omondi. Low-error, High-speed Approximation of the Sigmoid Function for Large FPGA Implementations. *Journal of Signal Processing Systems*, 52(2):137–151, 2008.

[78] J.E. Volder. The CORDIC trigonometric computing technique. *Electronic Computers, IRE Transactions on*, (3):330–334, 2009.

[79] JS Walther. A unified algorithm for elementary functions. In *Proceedings of the May 18-20, 1971, spring joint computer conference*, pages 379–385. ACM, 1971.

[80] Y.H. Hu and H.H.M. Chern. A novel implementation of CORDIC algorithm using backward angle recoding (BAR). *Computers, IEEE Transactions on*, 45(12):1370–1378, 2002.

[81] C. Alippi and G. Storti-Gajani. Simple approximation of sigmoidal functions: realistic design of digital neural networks capable of learning. In *Circuits and Systems, 1991., IEEE International Sympoisum on*, pages 1505–1508. IEEE, 2002.

BIBLIOGRAPHY

[82] H. Amin, KM Curtis, and BR Hayes-Gill. Piecewise linear approximation applied to nonlinear function of a neural network. In *Circuits, Devices and Systems, IEE Proceedings-*, volume 144, pages 313–317. IET, 2002.

[83] M. Zhang, S. Vassiliadis, and JG Delgado-Frias. Sigmoid generators for neural computing using piecewise approximations. *Computers, IEEE Transactions on*, 45(9):1045–1049, 2002.

[84] Souani C. Torki K. Faiedh, H. and K. Besbes. Digital hardware implementation of neural system used for nonlinear adaptive prediction. *Journal of Computer Science*, 2(4):355–362, 2006.

[85] A. Armato, L. Fanucci, G. Pioggia, and D. De Rossi. Low-error approximation of artificial neuron sigmoid function and its derivative. *Electronics letters*, 45(21):1082–1084, 2009.

[86] C.W. Lin and J.S. Wang. A digital circuit design of hyperbolic tangent sigmoid function for neural networks. In *Circuits and Systems, 2008. ISCAS 2008. IEEE International Symposium on*, pages 856–859. IEEE, 2008.

[87] *High Speed VLSI Implementation of the Hyperbolic Tangent Sigmoid Function*, volume 1, 2008.

[88] C. Castelnuovo. Calcolo delle probabilità, v. 1, 1965.

[89] O. Mncer and W. Luk. Parameterized high throughput function evaluation for FPGAs. *The Journal of VLSI Signal Processing*, 36(1):17–25, 2004.

[90] S. Haykin. *Neural networks: a comprehensive foundation*. Prentice Hall PTR Upper Saddle River, NJ, USA, 1994.

[91] X.U. Guide. XtremeDSP Design Considerations. 1, 2004.

[92] L. Sekanina and S. Friedl. On routine implementation of virtual evolvable devices using COMBO6. In *Evolvable Hardware, 2004. Proceedings. 2004 NASA/DoD Conference on*, pages 63–70. IEEE, 2004.

[93] D.A. Robinson. A Method of Measuring Eye Movemnent Using a Scieral Search Coil in a Magnetic Field. *Bio-medical Electronics, IEEE Transactions on*, 10(4):137–145, 2008.

[94] L.R. Young and D. Sheena. Survey of eye movement recording methods. *Behavior Research Methods & Instrumentation*, 1975.

[95] A.O. DiScenna, V. Das, A.Z. Zivotofsky, S.H. Seidman, and R.J. Leigh. Evaluation of a video tracking device for measurement of horizontal and vertical eye rotations during locomotion. *Journal of neuroscience methods*, 58(1-2):89–94, 1995.

[96] A. Meyer, M. B
"ohme, T. Martinetz, and E. Barth. A single-camera remote eye tracker. *Perception and Interactive Technologies*, pages 208–211, 2006.

[97] Z. Zhu and Q. Ji. Novel eye gaze tracking techniques under natural head movement. *Biomedical Engineering, IEEE Transactions on*, 54(12):2246–2260, 2007.

[98] M.F. Land and DN Lee. Where we look when we steer. *Nature*, 369(6483):742–744, 1994.

[99] J.B. Pelz, R. Canosa, and J. Babcock. Extended tasks elicit complex eye movement patterns. In *Proceedings of the 2000 symposium on Eye tracking research & applications*, page 43. ACM, 2000.

[100] C.A. Rothkopf and J.B. Pelz. Head movement estimation for wearable eye tracker. In *Proceedings of the 2004 symposium on Eye tracking research & applications*, page 130. ACM, 2004.

[101] J.S. Babcock, J.B. Pelz, and J. Peak. The wearable eyetracker: a tool for the study of high-level visual tasks. In *Proceedings of the Military Sensing Symposia Specialty Group on Camouflage, Concealment, and Deception, Tucson, Arizona*. Citeseer, 2004.

[102] C.H. Morimoto, A. Amir, and M. Flickner. Detecting eye position and gaze from a single camera and 2 light sources. *Pattern Recognition*, 4:40314, 2002.

[103] C.H. Morimoto and M.R.M. Mimica. Eye gaze tracking techniques for interactive applications. *Computer Vision and Image Understanding*, 98(1):4–24, 2005.

[104] Lc.Technologies. Also available as `http://www.eyegaze.com`, 2010.

[105] Applied science laboratories Anon. Also available as `http://http//www.asl.com`, 2010.

[106] D. Li, J. Babcock, and D.J. Parkhurst. openEyes: a low-cost head-mounted eye-tracking solution. In *Proceedings of the 2006 symposium on Eye tracking research & applications*, pages 95–100. ACM, 2006.

[107] L. Dongheng, D. Winfield, and DJ Parkhurst. Starburst: A hybrid algorithm for video-based eye tracking combining feature-based and model-based approaches. In *Proceedings of the 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition.*, 2005.

[108] D.A. Forsyth and J. Ponce. *Computer vision: a modern approach*. Prentice Hall Professional Technical Reference, 2002.

[109] J.S. Babcock and J.B. Pelz. Building a lightweight eyetracking headgear. In *Proceedings of the 2004 symposium on Eye tracking research & applications*, pages 109–114. ACM, 2004.

[110] F. Mulvey, A. Villanueva, D. Sliney, R. Lange, S. Cotmore, and M. Donegan. Exploration of safety issues in eyetracking. *COGAIN Deliverable 5.4*, 2008.

[111] D. Sliney, D. Aron-Rosa, F. DeLori, F. Fankhauser, R. Landry, M. Mainster, J. Marshall, B. Rassow, B. Stuck, S. Trokel, et al. Adjustment of guidelines for exposure of the eye to optical radiation from ocular instruments: statement from a task group of the International Commission on Non-Ionizing Radiation Protection (ICNIRP). *Applied optics*, 44(11):2162–2176, 2005.

[112] W.J. Ryan, A.T. Duchowski, and S.T. Birchfield. Limbus/pupil switching for wearable eye tracking under variable

lighting conditions. In *Proceedings of the 2008 symposium on Eye tracking research & applications*, pages 61–64. ACM, 2008.

[113] J.M. Franchak, K.S. Kretch, K.C. Soska, J.S. Babcock, and K.E. Adolph. Head mounted eye tracking of infants' natural interactions: a new method. In *Proceedings of the 2010 Symposium on Eye Tracking Research Applications*, pages 21–27. Citeseer, 2010.

[114] JD Crawford and T. Vilis. Axes of eye rotation and Listing's law during rotations of the head. *Journal of neurophysiology*, 65(3), 1991.

[115] R.S. Allison, M. Eizenman, and B.S.K. Cheung. Combined head and eye tracking system for dynamic testing of the vestibular system. *Biomedical Engineering, IEEE Transactions on*, 43(11):1073–1082, 2002.

[116] W. Chen, M.J. Er, and S. Wu. Illumination compensation and normalization for robust face recognition using discrete cosine transform in logarithm domain. *Systems, Man, and Cybernetics, Part B: Cybernetics, IEEE Transactions on*, 36(2):458–466, 2006.

[117] S. Marsi and S. Saponara. Integrated video motion estimator with Retinex-like pre-processing for robust motion analysis in automotive scenarios: algorithmic and real-time architecture design. *Journal of Real-Time Image Processing*, pages 1–15, 2010.

[118] SA Ranawade. Face Recognition and Verification Using Artificial Neural Network. *International Journal of Computer Applications IJCA*, 1(14):23–30, 2010.

[119] E.H. Land and J.J. McCann. Lightness and retinex theory. *Journal of the Optical society of America*, 61(1):1–11, 1971.

[120] D.J. Jobson, Z. Rahman, and G.A. Woodell. Properties and performance of a center/surround retinex. *Image Processing, IEEE Transactions on*, 6(3):451–462, 2002.

[121] D.J. Jobson, Z. Rahman, and G.A. Woodell. A multiscale retinex for bridging the gap between color images and the human observation of scenes. *Image Processing, IEEE Transactions on*, 6(7):965–976, 2002.

[122] H. Wang, S.Z. Li, and Y. Wang. Generalized quotient image. 2004.

[123] H. Wang, S.Z. Li, and Y. Wang. Face recognition under varying lighting conditions using self quotient image. In *Automatic Face and Gesture Recognition, 2004. Proceedings. Sixth IEEE International Conference on*, pages 819–824. IEEE, 2004.

[124] A. Shashua and T. Riklin-Raviv. The quotient image: Class-based re-rendering and recognition with varying illuminations. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 23(2):129–139, 2002.

[125] S. Du and R. Ward. Wavelet-based illumination normalization for face recognition. In *Image Processing, 2005. ICIP 2005. IEEE International Conference on*, volume 2. IEEE, 2005.

[126] V. Struc, J. Zibert, and N. Pavesic. Histogram remapping as a preprocessing step for robust face recognition. *WSEAS Transactions on Information Science and Applications*, 6(3):520–529, 2009.

[127] W.E.L. Grimson and D.P. Huttenlocher. On the sensitivity of the Hough transform for object recognition. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 12(3):255–274, 2002.

[128] N. Bennett, R. Burridge, and N. Saito. A method to detect and characterize ellipses using the Hough transform. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 21(7):652–657, 2002.

[129] A. Fitzgibbon, M. Pilu, and R.B. Fisher. Direct least square fitting of ellipses. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 21(5):476–480, 2002.

[130] E. Maini. Robust ellipse-specific fitting for real-time machine vision. *Brain, Vision, and Artificial Intelligence*, pages 318–327, 2005.

[131] G. Schiavone, D. Formica, F. Taffoni, D. Campolo, E. Guglielmelli, and F. Keller. Multimodal Ecological Technology: From Childs Social Behavior Assessment to Child-

Robot Interaction Improvement. *International Journal of Social Robotics*, 2010.

[132] H.W. Lilliefors. On the Kolmogorov-Smirnov test for normality with mean and variance unknown. *Journal of the American Statistical Association*, 62(318):399–402, 1967.

[133] W.H. Kruskal and W.A. Wallis. Use of ranks in one-criterion variance analysis. *Journal of the American statistical Association*, 47(260):583–621, 1952.

[134] G.R. Stavig and J.D. Gibbons. Comparing the mean and the median as measures of centrality. *International Statistical Review/Revue Internationale de Statistique*, 45(1):63–70, 1977.

[135] P. Ekman, R.W. Levenson, and W.V. Friesen. Autonomic nervous system activity distinguishes among emotions. *Science*, 1983.

[136] R.W. Levenson. Autonomic nervous system differences among emotions. *Psychological Science*, 3(1):23, 1992.

[137] J.L. Andreassi. *Psychophysiology Human behavior and physiological response*. Lawrence Erlbaum Assoc Inc, 2006.

[138] G.L. Lohse and E.J. Johnson. A comparison of two process tracing methods for choice tasks. In *System Sciences, 1996., Proceedings of the Twenty-Ninth Hawaii International Conference on,*, volume 4, pages 86–97. IEEE, 2002.

[139] J.P. Zbilut and C.L. Webber Jr. Recurrence quantification analysis.

BIBLIOGRAPHY

[140] T.M. Cover and P.E. hart. Nearest neighbor pattern classification. *IEEE Transactions on Information Theory*, 13(1):21–27, January 1967.

[141] R. Kohavi and F. Provost. Glossary of terms. *Machine Learning*, 30:271 274, 1998.

[142] N. Marwan, M. Carmen Romano, M. Thiel, and J. Kurths. Recurrence plots for the analysis of complex systems. *Physics Reports*, 438(5-6):237–329, 2007.